



The Next Wave

The National Security Agency's Review of Emerging Technologies
Vol 18 No 4 • 2011

The Art of Forecasting and Futures Planning

Will Carbon Be the New Silicon?

The Security Impact of System Virtualization

Thin Film/Miniature Power Sources





Letter from the Editor

Like every research organization, NSA's Research Directorate (RD) wants to know "What's the next big thing?" And, like every research organization, RD reviews current technology developments in the constant quest to predict "the next big thing" and to avoid technology surprise.

The Next Wave itself is part of this quest—our mission, after all, is to report on emerging technologies. In addition, RD acquires in-depth studies of technology topics to gauge global research and development efforts. Most of these studies are generated by Department of Energy national laboratories, which focus on cutting edge research and science-based solutions to hard national challenges.

This TNW issue merges the two efforts as we present several articles based on in-depth studies prepared by Sandia National Laboratories (Sandia) and Pacific Northwest National Laboratory (PNNL). The articles cover a broad range of topics including virtualization and security, graphene as a replacement for silicon, and thin film/miniatue power sources.

In addition to the serious scientific forecasts, our Departments Editor has written a lighter article about technology forecasting in general. He especially enjoyed researching technology forecasts from the Roman-era to the present day that turned out to be wrong. So, please enjoy.

The Next Wave is published to disseminate significant technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the U.S. Government. Articles present views of the authors and not necessarily those of NSA or the TNW staff.



For more information, please contact us at
TNW@tycho.ncsc.mil





CONTENTS

FEATURES

- 4 The Art of Forecasting and Futures Planning
 - 8 Will Carbon Be the New Silicon?
 - 18 The Security Impact of System Virtualization
 - 25 Thin Film/Miniature Power Sources
-

The Art of Forecasting and Futures Planning



"What can be more palpably absurd than the prospect held out of locomotives traveling twice as fast as stagecoaches?"

– *The Quarterly Review*, England, 1825

"This 'telephone' has too many shortcomings to be seriously considered as a means of communication. The device is inherently of no value to us."

– *A memo at Western Union*, 1876



1876

1883

"X-rays will prove to be a hoax."

– Lord Kelvin, President of the Royal Society, 1883



1838

"Men might as well project a voyage to the Moon as attempt to employ steam navigation against the stormy North Atlantic Ocean."

– Dr. Dionysius Lardner, professor of Natural Philosophy and Astronomy at University College, London, 1838

1825



"...so many centuries after the Creation it is unlikely that anyone could find hitherto unknown lands of any value."

– Advisors to King Ferdinand and Queen Isabella of Spain regarding a proposal by Christopher Columbus, 1486

1486

"I also lay aside all ideas of any new works or engines of war, the invention of which long-ago reached its limit, and in which I see no hope for further improvement."

– Roman engineer Julius Sextus Frontinus, cAD 84

cAD 84

"There is not the slightest indication that nuclear energy will ever be obtainable. It would mean that the atom would have to be shattered at will."

— Albert Einstein, 1932

"A rocket will never be able to leave the Earth's atmosphere."

— New York Times, 1936

2004

1949

1936

1932

1911

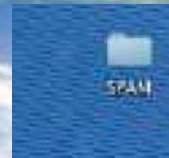


"Airplanes are interesting toys, but of no military value."

— Marshal Ferdinand Foch, future WWI Supreme Commander of the Allied Armies, 1911

"Where a calculator on the ENIAC is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1,000 vacuum tubes and weigh only 1.5 tons."

— Popular Mechanics, March 1949



"Two years from now, 'spam' will be solved."

— Microsoft chairman Bill Gates, 2004

Peering into the crystal ball can yield images that are cloudy and even upside-down. Still, everyone seems compelled to look. People are more than just curious to know what the future holds—we obsess over weather forecasts, stock market predictions, and the odds on the *big game*.

The art of futuring has been practiced in various forms over the course of human history. Today's forecasters, professionally called futurists or futurologists, may rely on statistical models run through massive supercomputers, but their quest to foresee the future is little different than it was for oracles, soothsayers, and prophets of yore. Although knowing what to expect might make it possible to change the course of human events, tempting fate can be a futile exercise, even in modern times. Often, the best that can be hoped

I predict...

"...four or five frigates will do the business without any military force."

— Lord North, British Prime Minister, debating imposing the Stamp Act on the American Colonies, 1774

"How, sir, would you make a ship sail against the wind and currents by lighting a bonfire under her deck? I pray you, excuse me, I have not the time to listen to such nonsense." — Napoleon Bonaparte, when told of Robert Fulton's steamboat, 1800s

"Rail travel at high speed is not possible because passengers, unable to breathe, would die of asphyxia." — Dr Dionysius Lardner, professor of Natural Philosophy and Astronomy, University College London, 1828

"It's a great invention but who would want to use it anyway?" — Rutherford B. Hayes, US President, after a demonstration of Alexander Bell's telephone, 1877

for is to glimpse the future and try to prepare for it.

On technology's horizon, numerous scenarios considered inevitable today were practically inconceivable a generation ago. Fuel cells, nanobots, unmanned vehicles, metamaterials, and telepathic communication are already finding their way into practical use. Time travel, invisibility cloaks, cyborgs, quantum computers, and human clones might not be all that far off.

Futuring helps us imagine the unimaginable and anticipate the unforeseen. In a world rapidly transformed by technology, the unimaginable can suddenly turn into crisis and the unforeseen the status quo. To prevent being taken by technological surprise, it is essential to anticipate the unexpected. Industry relies on technology forecasts to set production goals, order

commodities, and schedule expansion. For the intelligence community, the stakes are much higher.

A variety of futuring techniques are used to predict the world of tomorrow. Occasionally, a single event or discovery or even an idea changes everything. In those cases, human imagination is best suited to pierce the veil of the unknown.

Change, however, is usually incremental. A line drawn from the past and through the present can be extended into the future to point to a logical expectation. More deterministic methods of prognostication can be used to construct forecasts on the extrapolated data points of such trend lines. Trend analysis might miss the unexpected, but it provides a foundation for making decisions. If you know the local department store has held a “white sale” every Presidents Day for the past 10 years, you’re likely to put off buying that set of Egyptian cotton sheets until February.

With technology, no gauge for tracking a trend is more popularly watched than Moore’s Law. What started as a general observation about the pace of growth in computing power has come to set benchmarks for maintaining that pace. But even forecasts are fair game for forecasters to speculate on their sustainability. In the case of Moore’s Law, predictions about the trend in computing power coming to an end have been consistently proven wrong, however, with the original forecast trumping subsequent forecasts to the contrary.

Reading the tea leaves of statistical data is a tricky business that yields different conclusions depending on how the cup is tipped. Opinions among subject matter experts and respected pundits are often at odds. Divergent views can be useful for contingency planning, but to develop an overall strategy for

“When the Paris Exhibition closes, electric light will close with it and no more will be heard of it.” – *Erasmus Wilson, Oxford professor, 1878*

“The phonograph has no commercial value at all.” – *Thomas Edison, 1880s*

“We are probably nearing the limit of all we can know about astronomy.” – *Simon Newcomb, astronomer, 1888*

“Fooling around with alternating current is just a waste of time. Nobody will use it, ever.” – *Thomas Edison, 1889*

“Heavier-than-air flying machines are impossible.” – *Lord Kelvin, President of the Royal Society, 1895*

“Radio has no future.” – *Lord Kelvin, former President of the Royal Society, 1897*

“The ordinary ‘horseless carriage’ is at present a luxury for the wealthy; and although its price will probably fall in the future, it will never, of course, come into as common use as the bicycle.” – *Literary Digest, 1899*

“I must confess that my imagination refuses to see any sort of submarine doing anything but suffocating its crew and floundering at sea.” – *H.G. Wells, 1901*

“The horse is here to stay but the automobile is only a novelty—a fad.” – *The president of the Michigan Savings Bank advising Henry Ford’s lawyer, Horace Rackham, not to invest in the Ford Motor Company, 1903*

“I confess that in 1901, I said to my brother Orville that man would not fly for fifty years.... Ever since, I have distrusted myself and avoided all predictions.” – *Wilbur Wright, 1908*

“...the automobile has practically reached the limit of its development is suggested by the fact that during the past year no improvements of a radical nature have been introduced.” – *Scientific American, 1909*

“The coming of the wireless era will make war impossible, because it will make war ridiculous.” – *Guglielmo Marconi, 1912*

future development, a targeted approach depends on agreeing on what the future is most likely to look like.

Through visioning exercises organizations decide which potential future scenarios are most desirable, and they develop goals and the strategies to achieve them. Just as professional athletes *see* themselves powering through a fastball or executing a perfect dive, an enterprise can collectively envision the ideal customer experience, and an institution can foresee realizing its ultimate objectives. To this end, activities such as surveys and polls are used to arrive at a consensus. Consensus forecasting can be conducted as casually as by a roundtable discussion or through more formal techniques such as a methodical Delphi exercise.

For modern futurists new technologies and algorithms are providing innovative tools for generating more precise and reliable forecasts. Computer simulations are often used to model behaviors ranging from interactions among quantum dots to cosmic expansion after the Big Bang. Some of the largest supercomputers have been designed for just such purposes. The video game industry has also contributed to behavioral forecasting by providing platforms for experimenting with complex social and environmental interactions. All things considered, orcs in *World of Warcraft* don’t act that much different than travelers waiting in queues for airport security screening.

No matter how accurate, deterministic projections merely set the stage for scenario forecasts, painting the backdrop and furnishing a few basic props. Although statistical models might be useful for calculating actuarial tables, more random, or stochastic, futuring approaches are needed to animate a

“The cinema is little more than a fad. It’s canned drama. What audiences really want to see is flesh and blood on the stage.” – *Charlie Chaplin, 1916*

“The idea that cavalry will be replaced by these iron coaches is absurd. It is little short of treasonous.” – *Comment of Aide-de-camp to Field Marshal Haig at a tank demonstration, 1916*

“There is no likelihood man can ever tap the power of the atom.” – *Robert Millikan, physicist and Nobel Prize winner, 1923*

“The wireless music box has no imaginable commercial value. Who would pay for a message sent to no one in particular?” – *Associates of David Sarnoff responding to the latter’s call for investment in the radio in 1921*

“It is difficult to say what is impossible, for the dream of yesterday is the hope of today and the reality of tomorrow.” – *Robert Goddard, 1927*

“Who the hell wants to hear actors talk?” – *Harry Warner, Warner Bros., 1927*

“There will never be a bigger plane built.” – *A Boeing engineer, after the first flight of the 247, a twin engine plane that holds ten people, 1933*

“...any one who expects a source of power from the transformation of these atoms is talking moonshine...” – *Ernest Rutherford, physicist, 1933*

“Atomic energy might be as good as our present-day explosives, but it is unlikely to produce anything very much more dangerous.” – *Winston Churchill, First Lord of the Admiralty, then soon-to-be British Prime Minister, 1939*

“The name of Igor Sikorsky will be as well known as Henry Ford’s, for his helicopter will all but replace the horseless carriage as the new means of popular transportation. Instead of a car in every garage, there will be a helicopter...” – *Harry Bruno, aviation publicist, 1943*

complex scene. There is, perhaps, nothing more stochastic than human imagination.

Science fiction writers often provide the earliest and most vivid depictions of the future. The tomorrows imagined by visionaries such as Jules Verne and Robert Heinlein may have seemed incredible when they were first published, yet many of their predications have come true only decades later. And it took thousands of years of invention before the flight of Icarus was realized. Can we be certain that the age-old ambitions of turning lead to gold or reanimating a corpse are unobtainable?

Fiction is frequently steeped in the myths and legends of human history. A common and effective way to project the future is to filter it through the past. Historical analysis provides a forecasting method that benefits from hindsight to get a glimpse of tomorrow. Much can be learned by applying the lessons of history, as well as by reimagining momentous events in the past. By asking, “What if?” historical fiction can cast new light on current affairs that might prevent a repeat of prior mistakes.

Forecasting methods are as varied as the reasons for wanting to know “What’s next?” Futurists can apply their craft to sound a warning to prepare for impending change; they can have a hand in shaping the future, as well. Kings have turned to futurists to foretell their fates. Generals to predict success in battle. Adventurers to invoke good fortune. Modern forecasting methods may be based in scientific principles, but the desire to see into the future is as old as dreams.

“We are limited, not by our abilities, but by our vision.” – *author unknown* 📖

“[Television] won’t be able to hold on to any market it captures after the first six months. People will soon get tired of staring at a plywood box every night.” – *Darryl Zanuck, movie producer, 20th Century Fox, 1946*

“Television won’t last. It’s a flash in the pan.” – *Mary Somerville, pioneer of radio educational broadcasts, 1948*

“It would appear we have reached the limits of what it is possible to achieve with computer technology.” – *John von Neumann, computer scientist, 1949*

“It will be gone by June.” – *Variety, commenting on rock ‘n roll in 1955*

“Nuclear-powered vacuum cleaners will probably be a reality in 10 years.” – *Alex Lewyt, president of vacuum cleaner company Lewyt Corp., in the New York Times, 1955*

“Space travel is utter bilge.” – *Dr Richard van der Reit Wooley, Astronomer Royal, space advisor to the British government, 1956*

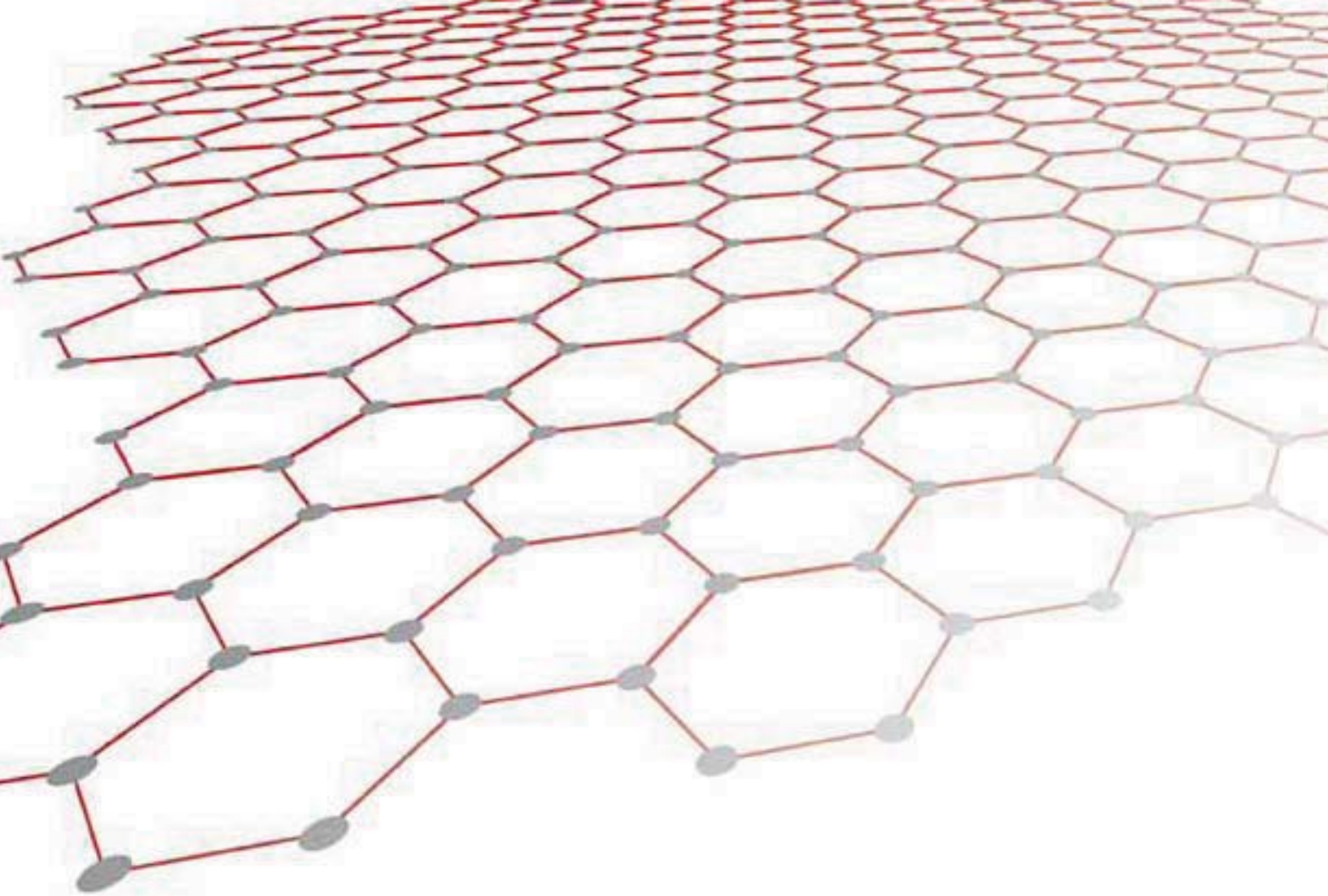
“The world potential market for copying machines is 5000 at most.” – *IBM, to the eventual founders of Xerox, 1959*

“Before man reaches the moon, your mail will be delivered within hours from New York to Australia by guided missiles. We stand on the threshold of rocket mail.” – *Arthur Summerfield, US Postmaster General, 1959*

“There is practically no chance communications space satellites will be used to provide better telephone, telegraph, television, or radio service inside the United States.” – *T. Craven, FCC Commissioner, 1961*

“It will be years—not in my time—before a woman will become Prime Minister.” – *Margaret Thatcher, October 26th, 1969*

“For all predictions do to this belong: That either they are right, or they are wrong.” – *John Tulley’s Almanac for 1688*



Will Carbon Be the New Silicon?

Pacific Northwest National Laboratory

Silicon, the eighth most common element in the universe[1] and the second most abundant element in the Earth's crust[2], is the mainstay of the electronics industry and the key component of most semiconductor devices, including computer chips.

Silicon has proved to be the most useful semiconductor constituent, remaining a semiconductor at higher temperatures than its chemical analog germanium, which is below it on the periodic chart. And silicon is less reactive than its other chemical analog, carbon, which is above it on the periodic chart. (See the section of the periodic chart in Figure 1, on page 10.) Our ability to manipulate silicon has steadily improved over the last several decades with the result that we continue to be able to build ever smaller electronic devices.

In a famous 1965 paper, Gordon Moore[3], then at Fairchild Semiconductor before helping to found Intel, observed from just four data points that “by 1975, the number of components per integrated circuit for minimum cost will be 65,000.” When Moore revisited the issue of exponential growth in 1975[4], he noted that a 16 kilobit charge-coupled device memory chip did indeed exist with about the number of transistors he predicted. This finding confirmed Moore’s forecast, although he lowered the slope of the growth curve he initially projected to a doubling period of every two years; concluding that circuit cleverness had run out of ideas—in other words, device layout had reached its optimal density. This doubling of transistors every two years has come to be known as Moore’s Law and the trend continues to this day. The 1965 paper noted that Moore was “one of the new breed of electronic engineers, schooled in the physical sciences rather than in electronics.” His PhD from CalTech was in physical chemistry. Material science is still driving progress in electronics.

Transistor density in processor chips, which is typically less than in memory chips, caught up four years later, in 1979, when Motorola introduced a comparably dense processor in the 32-bit MC68000 with 68,000 transistors. Tracking the number of transistors on a processor chip from 1971 (Intel 4004, 2,300 transistors[5]) to 2010 (NVIDIA GTX 480 aka Fermi, 3.0 billion transistors[6]), the doubling trend is remarkably consistent (Figure 2). Projecting the trend backwards gives 1950 as the year with one transistor; the transistor was actually invented in 1947[7]. If the trend continues, there will be about 66 billion transistors per chip in 2020 and 2.3 trillion transistors in 2030. The end

of Moore’s Law has been predicted many times, but so far a series of technological advances, all relying on silicon and its oxide, have managed to maintain the overall trend. How and whether the trend continues is an open question. The end of the current technology, called Complimentary Metal-Oxide Semiconductor or CMOS, may not be far off[8], but estimates based only on basic physical limits of computation have estimated that Moore’s Law could continue at its present rate for more than 600 years[9].

Until recently, making transistors smaller also made them run faster. This relationship was pointed out by Robert Dennard[10] at IBM in 1974 and is referred to as Dennard scaling. Furthermore, the energy reduction achieved by using smaller transistors exactly matches the energy increase from having more transistors. Using more transistors without reducing their energy requirements leads to the situation where not all the transistors on a chip can be energized because of the limited total power budget. The hard part, as you might suspect, is finding a way to make transistors smaller in a way consistent with high volume manufacturing while ensuring sufficient power to operate them. The scaling that Dennard foresaw did not really take place until the early nineties when the shift to CMOS was nearly complete.

To double the number of transistors in two years means the dimension of the transistor needs to be reduced by one over the square root of two, 0.71, or by half (the square of 0.71) every four years. For over forty years the semiconductor industry has succeeded in developing lithographic techniques to pattern smaller feature sizes, growing thinner gate oxides and reducing defect levels at increasingly challenging dimensions. The lithographic process is conceptually similar to chemical photography. Transistors are “developed” on a chip covered with a photosensitive material using multiple exposures of light projected through a series of masks. Moore and his colleagues started with 16-millimeter movie-camera lenses[11] and visible light. Now the process uses multi-million dollar lenses that weigh nearly half a metric ton and light deep in the ultraviolet, where the wavelength of light is shorter, to enable the patterning of smaller devices.

5 10.811 B Boron	6 12.011 C Carbon	7 14.007 N Nitrogen
13 26.982 Al Aluminum	14 28.086 Si Silicon	15 30.974 P Phosphorus
31 69.982 Ga Gallium	32 72.64 Ge Germanium	33 74.922 As Arsenic
Metal	Semimetal	Nonmetal

Figure 1: Silicon and surrounding elements on the periodic chart

Increasing lens size and reducing wavelength can't continue indefinitely, which means new techniques are called for to continue shrinking transistors.

The nineties saw exceptional improvements in processor clock speeds, going from the 40 MHz Motorola 68040, in 1990, to the 2 GHz Pentium 4, in 2000; a factor of 50 in just 10 years. During the decade, clock speed doubled every 21 months, besting the 24-month doubling rate for the number of transistors prescribed by Moore's Law and better than the historical average of clock speed doubling about every three years. In recent years, the pace of clock speed doubling has slowed to every four years or more, indicating that Dennard scaling may have reached an end. Dennard assumed that the threshold voltage, the minimum voltage needed for a transistor, would scale along with the operating voltage, providing both improved performance and power efficiency. This assumption held through the seventies when sub-threshold leakage was negligible on logic circuits, but now leakage is making it very difficult to scale operating voltage and hence performance. As the potential for reducing gate oxide thickness reaches an end, a growing fraction of the power put into a transistor simply heats the chip without making it switch faster. More worrying is that unless the power per transistor can be reduced there won't be enough power to operate all of them at once, a phenomenon referred to as dark silicon[12].

Intel's 65 nm generation transistors use a silicon dioxide dielectric only five atoms thick[13]. Efforts are now focused on using thicker materials with higher dielectric constants, the so-called high- κ materials[14], such as hafnium oxide (HfO), to minimize leakage. In 2010 Intel is using a 32 nm process[15] to produce the Westmere-EP processor. Furthermore, Dennard assumed that channel doping, which controls the energy needed for electrons to access the conduction band, could be kept appropriate for shorter channel lengths, but when channel doping gets too high the charge carrier mobility and performance degrade due to increased scattering from the impurities, and leakage increases due to quantum mechanical tunneling. The higher dielectric materials, which took over 10 years to go from laboratory prototype to commercial production, help minimize the leakage, but the fundamental problem remains. Furthermore, unlike transistors, scaled interconnects do not speed up. This means as the wires get smaller they don't keep up; the reduction in line capacitance is offset by an increase in resistance. This inverse relationship was not an issue in 1974, when interconnect delay was a small fraction of the clock cycle time, but it is now. Current strategies to reduce leakage include adding more metal layers (more wires); converting from aluminum to more conductive copper wires; and replacing SiO₂ dielectrics with superior dielectrics, such as HfO, to reduce capacitance.

The question now is whether or not silicon, its oxides, and its environment can be manipulated to achieve the size reduction necessary for Moore's Law to continue much longer. Are the processor performance improvements we have come to expect over? Is increased parallelism through the use of greater numbers of processor cores the only way to extend chip performance until transistors can't be made any smaller and Moore's Law really ends? Is the reign of silicon over? If so, the solution could come from carbon.

Carbon is the fourth most abundant element in the universe and, although not a significant component of the Earth's crust, it is the basis of all known forms of life. Carbon is very versatile and forms more compounds than any other element. The two most common pure forms of carbon are

diamond and graphite. The differences between the two are remarkable. Diamond is the hardest mineral known and graphite is one of the softest. Pure diamond is an insulator and a very excellent thermal conductor whereas graphite is an electrical conductor and can be used for thermal insulation. The differences can largely be attributed to the way carbon atoms bond to each other in the two forms. In diamond, every carbon atom is tightly bonded to four other carbon atoms in a tetrahedron using a type of covalent bonding called sp^3 , meaning the bonds are hybrids of one carbon 2S orbital and three 2P orbitals. This configuration forms four strong sigma type bonds. In graphite each carbon atom is tightly bonded to only three other carbon atoms in a hexagonal plane using a type of bonding called sp^2 , because one of the 2P orbitals forms a weaker pi bond instead of the stronger sigma bond. Graphite has no covalent or ionic bonds to carbons in other planes, only a very weak van der Waals bond resulting from fluctuating charge distributions. The weak bonding allows the slippage between planes, which makes graphite a good lubricant.

Both forms of carbon have been attracting the attention of the electronics industry. Boron doped diamond is a p-type semiconductor because boron has one fewer valence electron than carbon. The

thermal transport properties of diamond make it an attractive heat pipe to keep chips from overheating.

If graphite is thought of as a stack of hexagonal meshes like a pile of chicken wire, then an individual sheet of chicken wire would be graphene. It was long considered impossible for a single sheet of graphene to exist because thermal fluctuations normally force two-dimensional crystals to transform into more stable three-dimensional shapes. This misconception changed in 2004[16] when two Russian scientists working at the University of Manchester, Andrew Geim and his postdoc Konstantin Novoselov and six other co-workers managed to isolate a graphene sheet that was only one atom thick. They did this by using scotch tape to peel off layers of graphite until only one was left. What they found was a veritable wonder material. Geim and Novoselov were rewarded for their efforts with the 2010 Nobel Prize in Physics[17]. Free-standing graphene sheets are stable because they are not perfectly flat[18]. Intrinsic microscopic roughening from bends of several degrees and out-of-plane deformations that reach one nanometer gives the surface a corrugated effect. Graphene can also be grown on a substrate, which adds three-dimensional stability, and then lifted off as an intact two-dimensional crystal.

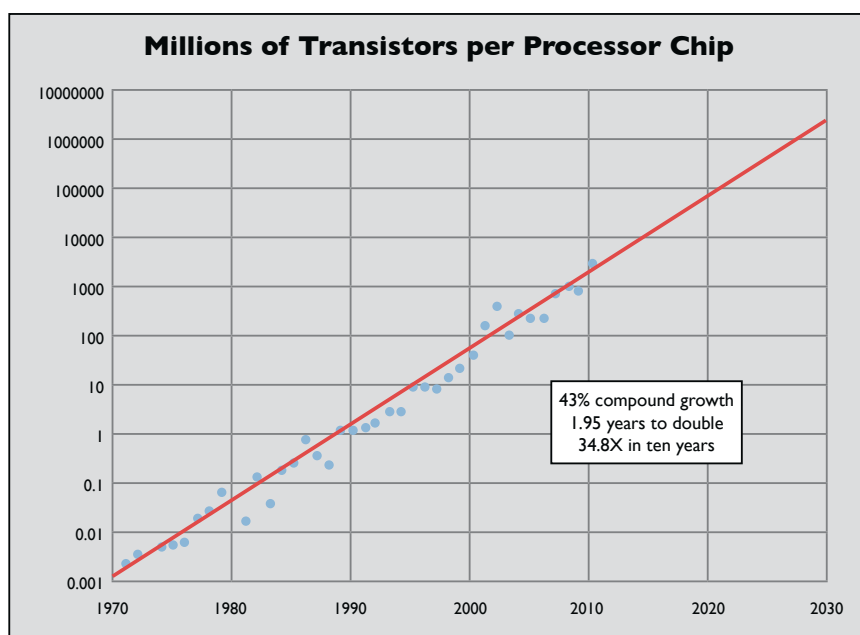


Figure 2: Exponential growth of the number of transistors per processor chip

Single sheets of graphene have been found in the graphite debris of pencil marks, once it became reasonable to look for them.

Graphene, being only one carbon atom thick and chemically stable, is not only the thinnest known material but, at 200 times the strength of steel[19], also the strongest. It can be stretched by 20 percent but is also very stiff under compression with a Young's modulus of about 1.0 TPa[20], which is comparable to diamond. Graphene is so robust and flexible that it can be folded; offering interesting possibilities for touch screens, bendable applications, and compact devices—picture a folded or scrolled piece of fabric containing integrated circuits. Graphene is also impermeable to gases, including helium. It is about twice as thermally conductive as diamond and 13 times more thermally conductive than copper. In addition, the thermal conductivity of graphene might be improved by changing its isotopic composition similar to the way it is in diamond—a diamond crystal that is 99.999 percent ^{12}C is 200 times more thermally conductive than naturally occurring diamond[21]. There doesn't appear to have been any attempts to create isotopically pure graphene—yet. And, unlike any other material, graphene shrinks with increasing temperature. Although graphene is stable well above room temperature, we don't know at what temperature it melts or even how it melts.

Most importantly, graphene is *unlike any other material electronically*.

Electrons moving through a graphene sheet behave more like photons than electrons in conventional materials. They can change their momentum and energy, but they cannot speed up or slow down. They lose their effective mass and become massless Dirac fermions, described by the relativistic Dirac equation rather than the Schrödinger equation that describes silicon. The electrons travel at a constant speed—a small (1/300th)[22] but noteworthy fraction of the speed of light and four times faster than in silicon. Furthermore, electrons in graphene can move ballistically—without collisions and the resultant heat—over great distances (micrometers) and through impurity barriers, even at room temperature. Graphene can conduct electrical current 10 to 100

times better than a silicon semiconductor at room temperature. The electron mobility of pristine graphene at room temperature in a laboratory has been measured at of over $200,000\text{ cm}^2/\text{V}\cdot\text{s}$ [23], ($>107\text{ cm}^2/\text{V}\cdot\text{s}$ using a bulk graphite substrate[24]) compared to silicon at around $1,400\text{ cm}^2/\text{V}\cdot\text{s}$. Early results from graphene films grown in a potential, but unoptimized, manufacturing environment have exhibited mobilities of up to $4,000\text{ cm}^2/\text{V}\cdot\text{s}$. Because the electrons are confined to a plane only one atom thick, they can be monitored and manipulated through nearby materials such as high-K materials, magnetic materials, superconductors, and molecules whose presence in small quantities are difficult to detect.

The high electron mobility of graphene has inspired the construction of field-effect transistors with steadily increasing cut-off frequencies: 26 GHz (December 2008)[25]; 100 GHz (February 2010)[26]; and 300 GHz (September 2010)[27]. These experiments are consistent with an eventual Terahertz device suitable for radiofrequency applications where a persistent off state is not required.

Building *logic* transistors, however, requires a semiconductor that can be turned on and off—and remain completely closed in the off state. Graphene sitting on a substrate of silicon dioxide is not a semiconductor, but a semimetal[28]. There is theoretical[29] evidence that suggests graphene suspended in a vacuum is actually a Mott[30] insulator. A Mott insulator does not insulate because the band is full, but because strong local electron correlations dynamically generate a charge gap by splitting a half-filled band into lower and upper Hubbard bands. If the band gap is large enough and can be controlled, it would make graphene suspended in free space a useful semiconductor. So far, however, such a state has not been demonstrated.

Areas of graphene can be made into useful semiconductors by controlling the substrate, adsorbate, or both, in other words, what is under or over the graphene sheet can turn graphene into a semiconductor. For example, hydrogen[31] will react with graphene to form graphane, which is a semiconductor with every other carbon atom pushed above or below the plane and with the

hydrogens sticking away from the plane. The reaction is reversible with the application of heat. In another example, theory[32] predicts that graphene deposited on a SiO₂ surface that has an oxygen-terminated surface will have a finite energy band gap. The band gap will be closed when the oxygen atoms are passivated with hydrogen atoms, which interfere with the way the oxygen lone-pair electrons interact with graphene. Graphene reacts with various atoms and molecules in a generally reversible way that usually leaves few defects behind. This resilience is because the underlying structure, the “chicken wire,” is usually only bent and not torn. Because graphene is essentially all surface and no bulk, the chemical surface effects tend to be exaggerated, leading to novel effects. Atoms or molecules on one surface can alter the energetics of the other side just one atom away, resulting in reactions that would be unstable if only one side was exposed. Graphene chemistry is currently in its infancy, but the prospects are bright.

Another approach to create suitable band gaps is to further constrain the electron wavefunction. Quantum confinement occurs when the wavefunction of a particle has about the same dimension as the space it occupies. This state occurs at the nanoscale with the consequences that the energy of the particle becomes discrete and the band gap is size dependent. With confinement the band gaps are likely to be larger than in the bulk material and more resistant to tunneling as a consequence. Tunneling can still occur, but the probability of success is normally proportional to the height of the barrier. This is not true of Klein[33] tunneling, where the probability actually increases if the barrier is above a height of $2mc^2$. The demonstration[34] of Klein tunneling in graphene is strong evidence that the charge carriers are indeed Dirac fermions.

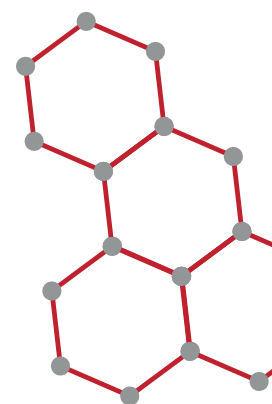
Narrow strips of graphene, nanoribbons, can become semiconductors under the right circumstances. The key is the orientation of the nanoribbon. Ones with more stable zig-zag edges are metallic, but nanoribbons cut from graphene rotated ninety degrees (as a hexagon, the atoms are sixty degrees apart) will have armchair edges and will be semiconductors[35]. It has been suggested that nanoribbons can conduct spin currents as well

and could possibly serve as the basis for nanoscale spintronic devices[36]. The width of the nanoribbon also influences the size of the band gap, with wider ribbons resorting to the behavior of the full graphene sheet. The controlled engineering of the nanoribbon edge structure is a significant challenge and will be required to obtain uniform performance of graphene-based devices[37].

The robustness of this quantum confinement means quantum effects are directly observable at room temperature, providing physicists with unprecedented access to a new quantum electrodynamics (QED) laboratory. Studies of graphene have led to new understandings of a raft of exotic quantum phenomena, including the half-integer quantum Hall effect, direct observations of Klein tunneling, Zitterbewegung, Schwinger production, supercritical atomic collapse, and Casimir-like interactions between absorbates on graphene. The electronic structure of graphene opens up many unexplored areas involving electro- and magneto-optics that could lead to efficient optoelectronic devices.

Alternative approaches involve multiple layers of graphene, “nanoroads[38]” of pristine graphene carved in the electrically insulating matrix of graphene sheets, and even graphite (ten or more layers of graphene) where a reproducible crack can be formed to create a large band gap for a nonvolatile memory device[39].

Role up a nanoribbon and you get a nanotube. A nanotube is a one-dimensional material whereas graphene is two-dimensional and bulk materials such as silicon are three-dimensional. Nanotubes differ from nanoribbons in that they are doubly degenerate because the wavefunction must be periodic around the circumference. Nanotubes are similar to nanoribbons in that their properties are determined by the orientation of the hexagon mesh and come in zig-zag, armchair, and “chiral” configurations, but armchair configurations are considered metallic and the others may only have very small to moderate band gaps. Mott insulating states have been identified in some normally metallic nanotubes[40]. Nanotubes also come in multi-walled varieties such as nested cylinders. Nanotubes have been the focus of much attention since the early



1990s[41]. Although working memory[42] and other devices[43] have been essentially hand-built in the laboratory using direct manipulation with an atomic force microscope (AFM), progress in volume manufacturing of consistent quality devices has been slow because of the difficulty of placing defect-free nanotubes of the desired configuration in the right place. Considerable effort has gone into investigating ways of producing high-quality nanotubes of the proper configuration, whether through direct growth controlled by catalysts or by separation of mixtures[44]. The research is driven not just by electronics, but by fields as diverse as medicine, hydrogen storage, desalination, composite materials, and solar power.


Volume manufacturing has been the real secret of silicon's success. Even if carbon, whether as graphene, nanotubes, graphite, or diamond, has superior properties at the nanoscale, without the ability to manufacture in volume it will remain an expensive, special-purpose or laboratory technology. Volume manufacturing at the nanoscale is a problem that confronts all technologies however, including silicon-based electronics. Conventional lithographic technology does not appear capable of reaching the nanoscale because the required wavelength is too small. Deep ultraviolet using 193-nanometer light is the most common current technology. Extreme ultraviolet light at 13.5 nm could be the next step, but that possibility faces many challenges (not the least of which is the energy of the photon smashing into the surface being patterned) as do the use of X-rays and gamma rays. Electron beams and AFMs have the resolution but are too slow except for building devices in the laboratory. Imprint lithography is not limited by wavelength, but has problems with defects and has yet to break out of the laboratory. Another laboratory effort that has possibilities is to build self-assembled structures using block copolymers.

Even further out are technologies based on metamaterials and nonlinear optics. Metamaterials are artificially designed composites possessing extraordinary optical properties that allow them to alter the propagation of electromagnetic waves, resulting in negative refraction, subwavelength imaging, and cloaking[45]. A metamaterial

superlens capable of subwavelength imaging could revolutionize the semiconductor lithographic process and provide a commercial path to nanoscale devices. Imaging directly through a nonlinear medium is currently not possible because intensity-dependent phase changes distort the image as the wave propagates through the material, but numerically reconstructing[46] the wave dynamics could lead to subwavelength lithography where the mask image is first scrambled by a computer before being projected onto the chip. Both these approaches are very computationally intensive.

The physics of real carbon-based devices have turned out to be much more complicated than can be explained by current theoretical interpretations. This doesn't mean the theory is wrong, just that our ability to make the approximations necessary to get an answer is inadequate. Harnessing these exotic effects is going to be challenging, but the result could be the emergence of nanoelectronic techniques that will continue to sustain Moore's Law. Will it be possible to grow large, folded, multilayer graphene sheets, where the unmodified areas are the high-speed interconnects and heat pipes and nanoscale size areas are modified to become semiconducting regions that form high-speed transistors or efficient optical-electronic interfaces? While there is reason for optimism, the field of prognostication is littered with the unfulfilled ideas of those who have predicted the end of silicon's dominance before. Recent experiments at Rice University[47] have resulted in a new way of building fast, robust, nonvolatile resistive switches and memories by making silicon nanocrystals 5 nm in diameter in silicon oxide—in the lab.

About Pacific National Laboratory

Pacific National Laboratory is a US Department of Energy national laboratory where interdisciplinary teams advance science and technology and deliver solutions to America's most intractable problems in energy, the environment, and national security. PNNL provides world-renowned scientists and engineers, facilities, and unique scientific equipment to strengthen US scientific foundations and advance scientific discovery through innovation. 

References and further reading:

- [1] Povh B, Rith K, Scholz C, Zetsche F. *Particles and Nuclei, An Introduction to the Physical Concepts*. 6th ed. Lavelle M, translator. Berlin: Springer-Verlag; 2008. p. 16. ISBN: 978-3-540-79367-0
- [2] Voronkov MG. Silicon Era. *Russian J. Appl. Chem.* 2007;80(12):2190-2196. Available from: doi:10.1134/S1070427207120397
- [3] Moore GE. Cramming more components onto integrated circuits. *Electronics*. 1965;38(8):114-117. Available from: doi:10.1109/JPROC.1998.658762
- [4] Moore GE. Progress in digital integrated electronics. *Electron Devices Meeting, 1975 International*. 1975 Dec 1-3;21:11-13. Available from: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1478174
- [5] The Intel 4004: Fun Facts [Internet]. Available from: <http://www.intel.com/about/companyinfo/museum/exhibits/4004/facts.htm>
- [6] NVIDIA's next generation CUDA compute architecture: Fermi. NVIDIA Whitepaper [Internet]. 2009 Oct 2. Available from: http://www.nvidia.com/content/PDF/fermi_white_papers/NVIDIA_Fermi_Compute_Architecture_Whitepaper.pdf
- [7] Brinkman WF, Haggan DE, Troutman WW. A History of the invention of the transistor and where it will lead us. *IEEE J. Solid-State Circuits*. 1997;32(12):1858-1865. Available from: doi:10.1109/4.643644
- [8] Zhirnov VV, Cavin RK III, Hutchby JA, Gourianoff GI. Limits to binary logic switch scaling - a gedanken model. *Proceedings of the IEEE*. 2003;91(11):1934-1939. Available from: doi:10.1109/JPROC.2003.818324
- [9] Krauss LM, Starkman GD. Universal limits on computation. 2004 May 10. Available from: arXiv:astro-ph/0404510v2
- [10] Dennard RH, Gaenssien RH, Rideout VL, Bassous E, LeBlanc AR. Design of ion-implanted MOSFETs with very small physical dimensions. *IEEE Journal of Solid State Circuits*. 1974;9(5):256-268. Available from: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1050511&isnumber=22538>
- [11] Arnold B. Shrinking possibilities. *IEEE Spectrum*. 2009;46(4):26-28, 50-56(NA). Available from: doi:10.1109/MSPEC.2009.4808761
- [12] Goulding N, Sampson J, Venkatesh G, Garcia S, Auricchio J, Babb J, Taylor MB, Swanson S. GreenDroid: a mobile application processor for silicon's dark future. *Proceedings of HotChips 22*; presented 2010 Aug 23. Available from: <http://cseweb.ucsd.edu/users/swanson/papers/HotChips2010GreenDroid.pdf>
- [13] Silicon innovation: Leaping from 90 nm to 65 nm. Intel Whitepaper [Internet]. 2006 March 6. Available from: http://www.google.com/url?sa=t&source=web&cd=1&ved=OCBgQFjAA&url=ftp%3A%2F%2Fdownload.intel.com%2Ftechnology%2Fsilicon%2Fsilicon_paper_06.pdf&rct=j&q=five%20atoms%20thick%2065%20nm%20intel&ei=tl3PTMTUOpGisQP0u7mjAw&sg=AFQjCNHpdNgQq98Tots5rbRElIOAQ9_v4Q
- [14] Wilk GD, Wallace RM, Anthony JM. High-K gate dielectrics: Current status and materials properties considerations. *J. Appl. Phys.* 2001;89(10):5243-5276. Available from: doi:10.1063/1.1361065
- [15] Introduction to Intel's 32nm process technology. Intel Whitepaper [Internet]. 2009 Dec 15. Available from: ftp://download.intel.com/pressroom/kits/32nm/westmere/Intel_32nm_Overview.pdf
- [16] Novoselov KS, Geim AK, Morozov SV, Jiang D, Zhang Y, Dubonos SV, Grigorieva IV, Firsov AA. Electric field effect in atomically thin carbon films. *Science*. 2004;306(5696):666-669. Available from: doi:10.1126/science.1102896
- [17] Press Release [Internet]. The Royal Swedish Academy of Science. 2010 Oct 05. Available from: http://nobelprize.org/nobel_prizes/physics/laureates/2010/press.html
- [18] Meyer JC, Geim AK, Katsnelson MI, Novoselov KS, Booth TJ, Roth S. The structure of suspended graphene sheets. *Nature*. 2007;446(7131):60-63. Available from: doi:10.1038/nature05545
- [19] Schewe P. World's strongest material. *Inside Science News Service* [Internet]. 2008 July 28; no. 867. Available from: <http://www.aip.org/pnu/2008/split/867-2.html>; Lee C, Wei X, Kysar JW, Hone J. Measurements of the elastic properties and intrinsic strength of monolayer graphene. *Science*. 2008;321(5888):385-388. Available from: doi:10.1126/science.1157996
- [20] Geim AK. Graphene: Status and prospects. *Science*. 2009;324(5934):1530-1534. Available from: doi:10.1126/science.1158877
- [21] Wei L, Kuo PK, Thomas RL. Thermal conductivity of isotopically modified single crystal diamond. *Phys. Rev. Lett.* 1993;70(24):3764-3767. Available from: doi:10.1103/PhysRevLett.70.3764
- [22] Castro Neto AH, Guinea F, Peres NMR, Novoselov KS, Geim AK. The electronic properties of graphene. *Rev. Mod. Phys.* 2009;81(1):109-162. Available from: doi:10.1103/RevModPhys.81.109
- [23] Bolotin KI, Sikes KJ, Jiang Z, Klima M, Fudenberg G, Hone J, Kim P, Stormer HL. Ultrahigh electron mobility in suspended

graphene. Solid State Communications. 2008;146(9-10):351-355. Available from: doi:10.1016/j.ssc.2008.02.024

[24] Neugebauer P, Orlita M, Faugeras C, Barra A-L, Potemski M. How perfect can graphene be? Phys. Rev. Lett. 2009;103(13):136403, 4pp. Available from: doi:10.1103/PhysRevLett.103.136403

[25] Lin Y-M, Jenkins KA, Valdes-Garcia A, Small JP, Farmer DB, Avouris P. Operation of graphene transistors at GHz frequencies. Nano Lett. 2008;9(1):422-426. Available from: doi:10.1021/nl803316h

[26] Lin Y-M, Dimitrakopoulos C, Jenkins KA, Farmer DB, Chiu H-Y, Grill A, Avouris Ph. 100-GHz transistors from wafer-scale epitaxial graphene. Science. 2010;327(5966):662. Available from: doi:10.1126/science.1184289

[27] Liao L, Lin Y-C, Bao M, Cheng R, Bai J, Liu Y, Qu Y, Wang KL, Huang Y, Duan X. High-speed graphene transistors with a self-aligned nanowire gate. Nature. 2010;467(7313):305-308. Available from: doi:10.1038/nature09405

[28] Wallace PR. The band theory of graphite. Phys. Rev. 1947;71(9):622-634. Available from: doi:10.1103/PhysRev.71.622

[29] Drut JE, Lähde TA. Lattice field theory simulations of graphene. Phys. Rev. B. 2009;79(16):165425, 14pp. Available from: doi:10.1103/PhysRevB.79.165425

[30] Phillips P. Mottness. Annals of Physics. 2006;321(7):1634-1650. Available from: doi:10.1016/j.aop.2006.04.003

[31] Elias DC, Nair RR, Mohiuddin TMG, Morozov SV, Blake P, Halsall MP, Ferrari AC, Boukhvalov DW, Katsnelson MI, Geim AK, Novoselov KS. Control of graphene's properties by reversible hydrogenation: Evidence for graphane. Science. 2009;323(5914):610-613. Available from: doi:10.1126/science.1167130

[32] Shemella P, Nayak SK. Electronic structure and band-gap modulation of graphene via substrate surface chemistry. Appl. Phys. Lett. 2009;94(3):032101, 3pp. Available from: doi:10.1063/1.3070238

[33] Klein O. Die Reflexion von Elektronen an einem Potentialsprung nach der relativistischen Dynamik von Dirac. Z. Phys. 1929;53(3-4):157-165. Available from: doi:10.1007/BF01339716

[34] Stander N, Huard B, Goldhaber-Gordon D. Evidence for Klein tunneling in Graphene p-n Junctions. Phys. Rev. Lett. 2009;102(2):026807, 4pp. Available from: doi:10.1103/PhysRevLett.102.026807

[35] Fujita M, Wakabayashi K, Nakada K, Kusakabe K. Peculiar localized state at zigzag graphite edge. J. Phys. Soc.

Jpn. 1996;65(7):1920-1923. Available from: doi:10.1143/JPSJ.65.1920

[36] Kusakabe K. Possible nano-spintronics devices with graphene as electron wave guides. 2009 APS March Meeting, 2009 March 16-20, Pittsburgh. Available from <http://meetings.aps.org/link/BAPS.2009.MAR.C1.166>

[37] Ritter RA, Lyding JW. The influence of edge structure on the electronic properties of graphene quantum dots and nanoribbons. Nature Materials. 2009;8(3):235-242. Available from: doi:10.1038/nmat2378

[38] Singh AK, Yakobson BI. Electronics and magnetism of patterned graphene nanoroads. Nano Lett. 2009;9(4):1540-1543. Available from: doi:10.1021/nl803622c

[39] Sinitskii A, Tour JM. Lithographic graphitic memories. ACSNano. 2009;3(9):2760-2766. Available from: doi:10.1021/nn9006225 [note: this turned out to be an artifact of the SiO_x substrate, reference 47]

[40] Deshpande VV, Chandra B, Caldwell R, Novikov DS, Hone J, Bockrath M. Mott insulating state in ultraclean carbon nanotubes. Science. 2009;323(5910):106-110. Available from: doi:10.1126/science.1165799

[41] Monthieux M, Kuznetsov VL. Who should be given the credit for the discovery of carbon nanotubes? Carbon. 2006;44(9):1621-1623. Available from: doi:10.1016/j.carbon.2006.03.019

[42] Rinkio M, Johansson A, Paraoanu GS, Törmä P. High-speed memory from carbon nanotube field-effect transistors with high-gate dielectric. Nano Lett. 2009;9(2):643-647. Available from: doi:10.1021/nl8029916

[43] Zavodchikova MY, Johansson A, Rinkio M, Toppari JJ, Nasibulin AG, Kauppinen EI, Törmä P. "Fabrication of carbon nanotube-based field-effect transistors for studies of their memory effects. physica status solidi (b). 2007;244(11):4188-4192. Available from: doi:10.1002/pssb.200776187

[44] a. Krupke R, Hennrich F, von Lohneysen H, Kappes MM. Separation of metallic from semiconducting single-walled carbon nanotubes. Science. 2003;301(5631):344-347. Available from: doi:10.1126/science.1086534

b. Chattopadhyay D, Galeska L, Papadimitrakopoulos F. A route for bulk separation of semiconducting from metallic single-wall carbon nanotubes. J. Am. Chem. Soc. 2003;125(11):3370. Available from: doi:10.1021/ja028599i

c. Chen ZH, Du X, Du MH, Rancken CD, Cheng HP, Rinzler AG. Nano Lett. 2003;3(9):1245-1249. Available from: doi:10.1021/nl0344763;

d. Arnold MS, Green AA, Hulvat JF, Stupp SI, Hersam MC. Sorting carbon nanotubes by electronic structure using density

- differentiation. *Nat. Nanotechnol.* 2006;1(1):60-65. Available from: doi:10.1038/nnano.2006.52;
- e. Collins PG, Arnold MS, Avouris P. Engineering carbon nanotubes and nanotube circuits using electrical breakdown. *Science.* 2001;292(5517):706-709. Available from: doi:10.1126/science.1058782;
- f. Zhang G, Qi P, Wang X, Lu Y, Li X, Tu R, Bangsaruntip S, Mann D, Zhang L, Dai H. Selective etching of metallic carbon nanotubes by gas-phase reaction. *Science.* 2006;314(5801):974-977. Available from: doi:10.1126/science.1133781;
- g. LeMieux MC, Roberts M, Barman S, Jin YW, Kim JM, Bao Z. Self-sorted, aligned nanotube networks for thin-film transistors. *Science.* 2008;321(5885):101-104. Available from: doi:10.1126/science.1156588;
- h. Dyke CA, Stewart MP, Tour JM. Separation of single-walled carbon nanotubes on silica gel. Materials morphology and raman excitation wavelength affect data interpretation. *J. Am. Chem. Soc.* 2005;127(12):4497-4509. Available from: doi:10.1021/ja042828h;
- i. Wang B, Poa CHP, Wei L, Li L-J, Yang Y, Chen Y. (n,m) Selectivity of single-walled carbon nanotubes by different carbon precursors on Co-Mo catalysts. *J. Am. Chem. Soc.* 2007;129(29):9014-9019. Available from: doi:10.1021/ja070808k;
- j. Bachilo SM, Balzano L, Herrera JE, Pompeo F, Resasco DE, Weisman RB. Narrow (n,m)-distribution of single-walled carbon nanotubes grown using a solid supported catalyst. *J. Am. Chem. Soc.* 2003;125(37):11186-11187. Available from: doi:10.1021/ja036622c;
- k. Li X, Tu X, Zaric S, Welsher K, Seo WS, Zhao W, Dai H. Selective synthesis combined with chemical separation of single-walled carbon nanotubes for chirality selection. *J. Am. Chem. Soc.* 2007;129(51):15770-15771. Available from: doi:10.1021/ja077886s;
- l. Li Y, Mann D, Rolandi M, Kim W, Ural A, Hung S, Javey A, Cao J, Wang D, Yenilmez E, Wang Q, Gibbons JF, Nishi Y, Dai H. Preferential growth of semiconducting single-walled carbon nanotubes by a plasma enhanced CVD method. *Nano Lett.* 2004;4(2):317-321. Available from: doi:10.1021/nl035097c;
- m. Qu L, Du F, Dai L. Preferential syntheses of semiconducting vertically aligned single-walled carbon nanotubes for direct use in FETs. *Nano Lett.* 2008;8(9):2682-2687. Available from: doi:10.1021/nl800967n;
- n. Li Y, Peng S, Mann D, Cao J, Tu R, Cho KJ, Dai H. On the origin of preferential growth of semiconducting single-walled carbon nanotubes. *J. Phys. Chem. B.* 2005;109(15):6968-6971. Available from: doi:10.1021/jp050868h;
- o. Ciuparu D, Chen Y, Lim S, Haller GL, Pfefferle L. Uniform-diameter single-walled carbon nanotubes catalytically grown in cobalt-incorporated MCM-41. *J. Phys. Chem. B.* 2003;108(2):503-507. Available from: doi:10.1021/jp036453i;
- p. Ural A, Li Y, Dai H. Electric-field-aligned growth of single-walled carbon nanotubes on surfaces. *Appl. Phys. Lett.* 2002;81(18):3464-3466. Available from: doi:10.1063/1.1518773;
- q. Huang S, Maynor B, Cai X, Liu J. Ultralong, well-aligned single-walled carbon nanotube architectures on surfaces. *Adv. Mat.* 2003;15(19):1651-1655. Available from: doi:10.1002/adma.200305203;
- r. Han S, Liu X, Zhou C. Template-free directional growth of single-walled carbon nanotubes on a- and r-plane sapphire. *J. Am. Chem. Soc.* 2005;127(15):5294-5295. Available from: doi:10.1021/ja042544x;
- s. Yuan D, Ding L, Chu H, Feng Y, McNicholas TP, Liu J. Horizontally aligned single-walled carbon nanotube on quartz from a large variety of metal catalysts. *Nano Lett.* 2008;8(8):2576-2579. Available from: doi:10.1021/nl801007r;
- t. Zhou W, Rutherglen C, Burke PJ. Wafer scale synthesis of dense aligned arrays of single-walled carbon nanotubes. *Nano Res.* 2008;1(2):158-165. Available from: doi:10.1007/s12274-008-8012-9;
- u. Kang SJ, Kocabas C, Ozel T, Shim M, Pimparkar N, Alam MA, Rotkin SV, Rogers JA. High-performance electronics using dense, perfectly aligned arrays of single-walled carbon nanotubes. *Nat. Nanotechnol.* 2007;2(4):230-236. Available from: doi:10.1038/nnano.2007.77;
- v. Ding L, Yuan D, Liu J. Growth of high-density parallel arrays of long single-walled carbon nanotubes on quartz substrates. *J. Am. Chem. Soc.* 2008;130(16):5428-5429. Available from: doi:10.1021/ja8006947;
- w. Ding L, Tselev A, Wang J, Yuan D, Chu H, McNicholas TP, Li Y, Liu J. Selective growth of well-aligned semiconducting single-walled carbon nanotubes. *Nano Lett.* 2009;9(2):800-805. Available from: doi:10.1021/nl803496s
- [45] Pendry J. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* 2000;85(18):3966-3969. Available from: doi:10.1103/PhysRevLett.85.3966
- [46] Barsi C, Wan W, Fleischer JW. Imaging through nonlinear media using digital holography. *Nature Photonics.* 2009;3(4):211-215. Available from: doi:10.1038/nphoton.2009.29
- [47] Yao J, Sun Z, Zhong L, Natelson D, Tour JM. Resistive switches and memories from silicon oxide. *Nano Lett.* 2010;10(10):4105-4110. Available from: doi:10.1021/nl102255r

The Security Impact of System Virtualization

Sandia National Laboratories

System virtualization has many benefits and is being adopted rapidly by businesses, governments, and individuals. Some of the biggest problems facing data center managers such as server sprawl, energy consumption, security, and rapid provisioning in response to changing load have solutions based on virtualization technology. To date, adoption has been broad but shallow. Most organizations have implemented some form of virtualization, but few have adopted it pervasively. As time goes on, more organizations will commit to a virtualized infrastructure including virtualized computation, communications, and storage.

Virtualization introduces several interesting system properties, such as isolation, inspection, interposition, and high availability, that have promising security applications. Isolation can be leveraged to develop systems that exhibit defense in depth, sandbox potentially malicious software, or separate mutually distrustful workloads while sharing infrastructure. Inspection allows software running at the virtualization layer to comprehensively and transparently observe the activity of a managed virtual machine for malicious behavior, misconfiguration, or performance anomalies, while interposition enables a hypervisor to take action against a misbehaving virtual machine or to implement new features such as a transparent VPN or intrusion prevention system. High availability features ease implementation of services that are resilient to hardware and certain kinds of software failures.



Background

System virtualization is a technique that allows a piece of software, usually called a virtual machine monitor (VMM) or hypervisor, to logically partition a single physical computer system into one or more virtual machines (VMs). Each virtual machine typically has a virtual hardware interface that is compatible with the original physical machine. Software designed to run on a comparable physical system will usually run unmodified in a virtual machine. A hypervisor is privileged system software, like an operating system, and uses many of the same techniques and depends on many of the same hardware features to accomplish its goals.

A virtual machine shares its physical host computer with the hypervisor and often with other virtual machines. A mechanism is required to prevent one virtual machine from tampering with other virtual machines or the hypervisor. The key idea underlying system virtualization is that CPU instructions, which are software's only access to the world, can be divided into safe and unsafe categories. Often, unsafe instructions are called sensitive instructions. By definition, safe instructions can be allowed to run without supervision, but sensitive instructions must be overseen and possibly modified or denied. The reason is that unsupervised sensitive

instructions could negatively affect other virtual machines or the hypervisor. Arithmetic operations, most control transfers, and most memory accesses are safe. In contrast, sensitive instructions change or reveal the configuration of the physical machine. Modifying the configuration of the memory management unit, interacting with I/O devices, and disabling interrupts are examples of sensitive operations.

A hypervisor uses a technique called depriving to ensure that software running within a virtual machine, usually called a guest, is allowed to run safe instructions unhindered, but that it will be notified of all sensitive instructions before they occur. The hypervisor does this by running all software within a virtual machine at a low privilege level that denies direct access to sensitive instructions. If a guest executes a sensitive instruction, the hardware raises an exception that transfers control to the hypervisor before the sensitive instruction takes effect. When notified that a sensitive instruction is about to occur, the hypervisor has the opportunity to supervise and modify its execution. The approach of detecting sensitive instructions and responding by safely emulating their behavior is called trap and emulate. This method is used by most high performance hypervisors to ensure that software running within a virtual machine remains confined there.

Trap and emulate requires that sensitive instructions consistently generate an exception when executed by unprivileged software. Until recently, this condition was not satisfied on IA-32, the most widely used microprocessor family for PC clients and servers. VMware developed novel software techniques to mitigate this limitation in the late 1990s, and, recently, IA-32 vendors such as Intel and AMD have added support for direct trap and emulate to their microprocessors. These developments are two important reasons system virtualization is widely available and so popular today.

The techniques used to implement system virtualization have been in continuous use on various platforms and have undergone refinement for several decades. Today, there is broad consensus about how to build a high performance, reliable



Figure 1: Logical organization of the software stack in a virtualized environment

hypervisor. The basic architecture of virtualization software is expected to remain about the same for the foreseeable future.

The introduction of the virtualization layer into the software stack has the potential, however, to act as a catalyst for change in adjacent system layers. For example, the hypervisor can naturally assume the role of hardware manager from the operating system, leaving the operating system free to concentrate on its role as the creator of convenient abstractions and programming environments for applications. By removing the burden of supporting the myriad hardware devices available on the market today, virtualization can lower the barrier to entry for new operating systems and encourage more diversity there, including operating systems tailored to specific applications.

Similarly, the existence of an independent virtualization software market seems to be an accident of the historically horizontal integration of the personal computing client and server industry. Most features of core virtualization fit comfortably within other systems layers, where a few powerful vendors dominate. Virtualization software could disappear as an independent entity and merge with adjacent layers like the operating system, the platform firmware, or both. In any case, it is safe to claim that widespread adoption of virtualization will apply pressure for the organization of the software stack to evolve in the near term.

Virtual machine isolation and security

A hypervisor implements a level of indirection between a virtual machine and the physical computer system on which it runs. That indirection makes it possible to draw a neat line around everything that is inside a virtual machine and clearly distinguish it from what is on the outside. By carefully managing the interface between the virtual machine and its environment, a hypervisor can prevent software running in a virtual machine from directly accessing the hypervisor, other virtual machines, or the rest of the outside world.

A hypervisor uses a variety of hardware and software mechanisms to isolate virtual machines. For example, it uses address translation provided

by the memory management unit to ensure that a virtual machine cannot reference memory that does not belong to it. The hypervisor schedules access to resources like the CPU and network interfaces to ensure that each virtual machine is allowed to use the resource without starving others of access. Separate virtual storage devices are provided to virtual machines whose names and address spaces do not overlap; hence one virtual machine has no means to refer to storage owned by another. Each of these techniques is possible because the hypervisor does not allow software in a virtual machine to directly execute any sensitive instructions, which could break isolation.

For many years, researchers, security engineers, and system administrators have exploited the isolation properties provided by a virtualized environment for a variety of purposes. For example, it is convenient to use virtual machines to create sandboxes in which malicious or buggy software can be tested and analyzed. If a virtual machine becomes infected or crashes, a previous, safe state can be restored quickly and easily. A virtual machine can be configured to ensure that attempts to infect other machines never leave the virtualized environment, reducing the need for physically isolated test networks and preventing accidental infestation. Virtual machines are being used increasingly to provide defense in depth by logically separating services like mail servers and web servers from one another even though they share a physical machine in a consolidated data center. Separating services limits the scope of a successful attack to a single vulnerable service rather than allowing it to spread to unrelated but incidentally collocated services. Finally, cloud computing service providers depend heavily on the isolation properties of virtual machines to maintain separation between workloads belonging to different customers, a form of multi-tenancy.

A great deal of effort has been spent on making virtual machine isolation strong while maintaining other important properties of virtualization such as performance and the ability to consolidate. Virtualization-based isolation depends critically on the integrity of the software interface between the virtual machine and the hypervisor. Traditionally,

the virtual machine interface has been considered stronger and less permeable than other software interfaces, such as the system call interface that separates user applications from an operating system. This confidence is based on the relatively smaller size of the hypervisor and the relative simplicity of the hardware-like interface it provides. These properties simplify analysis and provide fewer points of attack. So far, the virtual machine interface provided by many virtualization products has proven to be quite resistant to attack, but a small number of successful exploits have appeared publicly[1-3]. As is often true, most successful attacks do not target the strong core isolation of the hypervisor, but rather exploit softer, secondary services or convenience features.

Uncertainty about the actual strength of virtual machine isolation has slowed, prevented, or modified adoption of virtualization in some important application domains where security is important. For example, some cloud customers and vendors are, so far, unwilling to depend on virtualization to isolate sensitive information and computation belonging to different customers. Instead, they have opted to use physical partitioning or customer-aware virtual machine placement to mitigate their risks. Similar concerns exist in the malware analysis and multi-level security communities, each of which have adopted their own strategies to contain risk. There is nothing intrinsic to virtualization technology that would preclude nearly arbitrarily high levels of isolation. As of today, however, isolation is only one property of virtualized systems important to consumers, and the market for exceptionally high isolation is still small.

Virtual machine inspection and control

A VMM has the ability to inspect and control nearly every aspect of guest software execution. The content of guest memory and guest CPU registers can be read and written at hypervisor-controlled times. The guest program counter can be modified arbitrarily. Every I/O request can be intercepted, inspected, and modified. The VMM can arrange to be notified of a broad variety of guest events including instruction execution, access to specific

memory pages, function calls, and I/O events. In general, the VMM has complete and constant control of all architectural features of a guest virtual machine. Exceptionally powerful guest security analysis and protection features can be built on top of the access capabilities provided by a hypervisor. For example, anti-virus scanners, host and network intrusion detection systems, and system integrity monitors that operate at the virtualization layer have all been proposed or exist today.

Security services located within the virtualization layer are isolated from the potentially malicious software they are observing by the virtual machine interface. It is far more difficult for malware that deceives or disables security services to mount their attacks against hypervisor-based tools. Some research even claims that hypervisor-based security services can make themselves completely undetectable[4].

One significant weakness of hypervisor-based security analysis relative to guest-based analyses is that a hypervisor lacks detailed information about the semantics of the software running within a guest. For example, even though a VMM can read all memory associated with a virtual machine, there is no obvious way for it to know what processes are running within the guest and what their current state is without resorting to external assistance. Such information, however, is trivial to obtain from within the guest. This problem has been called the “semantic gap”[5], and a great deal of research has been done on how best to infer or otherwise obtain and use detailed and consistent information about operating system and application abstractions such as processes, threads, users, security tokens, network flows, and file system caches from within a hypervisor.

So far, there are a few promising approaches, but none has emerged as clearly superior in all cases. For example, one approach is to import detailed data structure definitions and semantic information from operating system debugging tools and use those to locate and understand operating system components of interest from within a hypervisor[6]. Another defines the information required by an analysis in terms of architectural information

naturally available to the hypervisor as part of its role in hardware virtualization[7]. The former leads to virtualization components that have significant insight into guest internals, but are tightly coupled to specific operating system vendors and versions and are brittle in the face of even minor operating system updates. The latter is less powerful, but is also much less dependent on the details of the guest and so is more broadly applicable. A third approach injects a software agent into the guest operating system with which a hypervisor can cooperate to extract and understand guest state. Commercial security products seem to prefer an agent-based approach because of its simple implementation, robustness, and broad applicability. Unfortunately, relying on an agent inside the guest negates many of the protection and transparency advantages of running within the virtualization layer.

The ability to transparently inspect software running within a guest has been available for many years in the form of full system emulators or simulators[8-9]. Emulators have often been used as debugging aids during the development of system software or for analyzing suspicious and potentially dangerous software in a security sandbox. What recent developments in system virtualization have added to this mix is the ability to perform transparent inspection and analysis at high speed on production systems. We can expect new detection, analysis, and remediation techniques that depend on emulator-like access and control, which would have been considered invasive and slow until recently, to appear on production systems based on virtualization technology.

Potential for new vulnerabilities

There are many explicit and implicit interfaces between untrusted guest software and the hypervisor including executing privileged instructions, interacting with emulated devices, and using virtualization-specific services like a dynamic host configuration protocol (DHCP) server on a virtualized network. Latent vulnerabilities within commercial hypervisors have been demonstrated that result in arbitrary code execution at the highest privilege on the host computer. Documented vulnerabilities of widely used virtualization software are rare today. Nevertheless, system

virtualization adds potentially vulnerable privileged interfaces to its guest environments.

In addition to the hypervisor itself, virtualization—especially enterprise deployments of virtualization—encompasses a whole ecology of management software, remote management APIs, and third-party components. Management software is used to create, start, stop, observe, debug, profile, archive, and otherwise control virtual machines. The management console wields enormous power over the virtualized enterprise. Enterprise management software for virtualized environments must be protected carefully to ensure whole-system security.

The virtualization layer, the applications based on it, and the management tools that control it contribute useful and compelling features to a virtualized system. It is not unlikely that the benefits of consolidation, transparent security services, and unified management outweigh the immediate security concerns over virtualization. It is clear, however, that the widespread deployment of virtualization represents an inflection point in the size and complexity of systems that can be built and managed. The gargantuan data centers of Google, Yahoo!, Facebook, and Amazon would have been unthinkable only a short time ago and have been enabled in part by virtualization technology. Increased complexity and deeper layering has historically been detrimental to system security.


Implications of virtualization on system security

Virtualization brings compelling features to individual computer systems and data centers. A significant fraction of server and corporate desktop computing will be performed within a virtualized environment over the next few years. It pays to examine how the introduction of virtualization technology changes the security environment. Proponents have argued that virtual machine isolation, high availability, and transparent, high speed introspection form the basis for a new level of security attainable via virtualization. More cautious observers have emphasized the unknown risks posed by introducing a new layer of privileged software, increasing network complexity, and proceeding before adopting accepted security practices for virtualized systems.

The true implications of virtualization on system security are still unclear. The hypervisor, which has mostly escaped the attention of malicious security research, is likely to become the target of more focused attacks and, as a result, will become more resilient. As virtualization is deployed, the mistakes made by early adopters will be recognized, corrected, and codified into a set of practices that prevent the most egregious configuration and policy errors. The growth of cloud computing will provide experience on how to effectively and safely share common infrastructure among mutually untrusting entities. In general, the level of uncertainty will be reduced, and virtualized computing will settle into the slightly uneasy security equilibrium we have today. As with many disruptive technology developments that initially hamper reliability and security, engineering practice catches up quickly

and the overall quality of systems typically improves. Hence, the true impact of virtualization on system security is likely neutral in the long run.

About Sandia National Laboratories

Sandia National Laboratories is a multi-program US Department of Energy national laboratory. Since 1949, Sandia National Laboratories has developed science-based technologies that support national security. Through science and technology, people, infrastructure, and partnerships, Sandia's mission is to meet national needs in five key areas: nuclear weapons, energy and infrastructure assurance, nonproliferation, defense systems and assessments, and homeland security and defense. 

References and further reading:

- [1] Kortchinsky K. Cloudburst: A VMware guest to host escape story [conference talk]. 2009 Black Hat. 25-30 July 2009; Las Vegas (NV). Slides available from: <http://www.blackhat.com/presentations/bh-usa-09/KORTCHINSKY/BHUSA09-Kortchinsky-Cloudburst-SLIDES.pdf>
 - [2] Mehta N, Smith R. VMWare DHCP server remote code execution vulnerabilities. IBM Internet Security Systems Protection Advisory [Internet]. Last updated 19 Sep 2007. Available from: <http://www.iss.net/threats/275.html>
 - [3] Shelton T. VMware NAT networking buffer overflow vulnerability. Secunia Advisory 18162 [Internet]. Last updated 27 Dec 2005. Available from: <http://secunia.com/advisories/18162/>
 - [4] Dinaburg A, Royal P, Sharif M, Lee W. Ether: Malware analysis via hardware virtualization extensions. In: Peng N, Syverson PF, Jha S, editors. CCS 2008. Proceedings of the 15th ACM Conference on Computer and Communications Security; 27-31 Oct 2008; Alexandria (VA). p. 51-62.
 - [5] Chen PM, Nobel BD. When virtual is better than real. In: HotOS-VIII. Proceedings of 8th Workshop on Hot Topics in Operating Systems; 20-23 May 2001; Elmau/Oberbayern, Germany.
 - [6] Garfinkel T, Rosenblum M. A virtual machine introspection based architecture for intrusion detection. In: NDSS 2003. Proceedings of the 10th Annual Network and Distributed System Security Symposium [Internet]; 6-7 Feb 2003; San Diego (CA). Available from: <http://www.isoc.org/isoc/conferences/ndss/03/proceedings/>
 - [7] Jones ST, Arpaci-Dusseau AC, Arpaci-Dusseau RH. Antfarm: Tracking processes in a virtual machine environment. In: USENIX '06. Proceedings of the USENIX 2006 Annual Technical Conference [Internet]; 1-3 June 2006; Boston (MA). Available from: <http://www.usenix.org/events/usenix06/tech/>
 - [8] Rosenblum M, Bugnion E, Devine S, Herrod SA. Using the SimOS machine simulator to study complex computer systems. ACM Transactions on Modeling and Computer Simulation (TOMACS). 1997;7(1):78-103.
 - [9] Magnusson PS, et al. Simics: A full system simulation platform. IEEE Computer. 2002;35(2):50-58.
- Sunbelt Software. CWSandbox [Internet]. Available from: <http://sunbeltsecurity.com/sandbox/>. [accessed 2010].

Thin Film/Miniature Power Sources

Pacific Northwest National Laboratory



Background and challenges

During the last two decades, advances in the miniaturization of electronic devices have greatly outpaced advances in battery technology[1]. Miniaturization of electronic devices follows a trend similar to that purported by Moore's Law, which describes a long-term trend in the history of computing hardware advances. The number of transistors that can be placed inexpensively on an integrated circuit has increased exponentially, doubling approximately every two years[2]. Fast processors continue to push the power and energy limit of onboard power supplies. Although significant effort has been made during the last two decades to improve the energy density of power sources, these efforts have led only to an improvement of about 15 percent per year in energy density, which is much less than the pace of Moore's Law. The reason for this technology lag is that power stored in a material

is proportional to its mass (or volume). High-speed microprocessors provide an increasing opportunity for advanced miniature applications such as microelectromechanical system (MEMS) devices, implantable miniature medical devices, and wireless sensor and actuator networks. However, the potential applications of microprocessors are significantly limited by the capabilities of the power sources required for their use.

Typical dimensions of power sources for microelectronic applications range from 10 cm (onboard backup power) to 10^{-2} cm (power for MEMS devices). The most critical parameter for these applications is the energy per unit volume (i.e., watt hours per liter—Wh/L) or energy per unit thickness (e.g., thin film power sources). Currently, most portable electronics use lithium (Li) or Li-ion batteries as their power sources. The energy density of a bulk Li-ion battery can be as high as ~500 Wh/L.



Recently developed Li-air batteries have a theoretical energy density of more than 3000 Wh/L. Primary Li-air batteries with a practical specific energy of up to 1200 Wh/L can be expected to be available within the next five years. Although rechargeable Li-air batteries also have attracted significant attention in recent years, there are still many barriers that need to be overcome before their application becomes practical.

All energy storage devices need to be appropriately packaged to extend their operation and shelf lives. One of the most significant challenges in miniature and thin film power sources is that the volume ratio of the active materials to the total volume (which includes the package) of miniaturized ($<1 \text{ mm}^3$) and thin film ($<0.3 \text{ mm}$ thick) batteries decreases rapidly as the battery size decreases. Reductions in the lateral dimension of a battery are limited by the accuracy of the manufacturing process. For mechanical masks, it is difficult to control the accuracy of the masks to less than 0.1 mm. The larger allowance required for quality control will lead to increased loss in device efficiency. For example, if the edge sealing needs to be 0.1 mm wide for a 1 mm square device, the area efficiency will be only 64 percent. More sophisticated size-control and alignment approaches, such as microlithography, will be needed before further reductions in the size of devices can be accomplished. The efficiency losses that accompany size reductions are common to all of the miniature devices. To reverse this trend and achieve high-efficiency power sources for miniature devices, a combination of technologies (including power scavenging, power storage, and power management) are needed to obtain the long-lasting power at the less-than-cubic millimeter scale.

State of the art in thin film/miniature power sources

Various miniature electronic devices require power sources with different characteristics. Some devices need high power pulses, while other devices require low-power, long-term discharge capabilities. Low-capacity batteries (i.e., $<1 \text{ mAh}$) are required for operation of MEMS devices or for backup power

for nonvolatile static random access memory. These devices normally need a power source with only a low energy density (i.e., $<10 \text{ Wh/L}$). Batteries with higher capacities are required to operate sensors, optical switches, smart cards, implantable medical devices, etc. These high-capacity batteries also often need to have higher energy densities.

Commercially available primary thin film batteries mainly include Li-MnO₂ and Zn-MnO₂ batteries. These paper-thin batteries are suitable for applications requiring low profile batteries where traditional button cell batteries are too thick. Most primary thin film batteries still use a polymer or gel electrolyte. One example is the “power-paper” cell made by Power Paper Ltd.^a, which is based on the printable Zn-MnO₂ system. The thickness of the power-paper cell is $\sim 0.6 \text{ mm}$. These batteries have been used in radio-frequency identification (RFID) tags, medical devices, electronic greeting cards, etc. Another example of a primary thin film battery is the Li-MnO₂ system manufactured by Solicore, Inc.^b. Solicore uses a semisolid electrolyte based on doped polyimide, and its batteries are $\sim 0.45 \text{ mm}$ thick. The batteries have been used in smart cards and for military applications such as the power supplies that can be attached to the collar of clothes. Many manufacturers have developed rechargeable polymer Li-ion batteries that are less than 0.5 mm thick. However, the energy densities of these cells decrease with decreasing thickness because the weight ratio of inactive materials in a battery increases with decreasing size. The typical thickness of the packaging material used for Li-ion batteries is $\sim 0.1 \text{ mm}$ thick (total thickness of the packaging material will be $\sim 0.2 \text{ mm}$ thick). Therefore, the energy density of thin film batteries will decrease dramatically when the thickness of the battery is close to the thickness of the packaging material. New packaging materials need to be developed to reduce the battery thickness to less than 0.2 mm.

The most successful rechargeable thin film batteries are based on a solid state, thin film electrolyte, Li phosphorus oxynitride (LiPON), which was developed by Bates et al.[3-5] at Oak Ridge National Laboratory (ORNL). The typical

^a www.powerpaper.com/?categoryId=43875

^b www.solicore.com/flexion-batteries.asp

structure of the thin film, solid state batteries is Li/LiPON/LiCoO₂. All of the components in these batteries are solid state inorganic materials prepared by sputtering or thermal evaporation. The total thickness of the active materials is normally less than 10 μm . The technology is ideally suited for a variety of applications where a compact but safe high-energy source is required. Nancy et al.^[6] demonstrated that test cells have maintained a charged state for more than a year with negligible loss of capacity. This level of reliability is not possible for other types of rechargeable batteries. The cycle life of these batteries can exceed more than 40,000 full-depth charge/discharge cycles, which far exceeds the cycle life of other types of batteries. There is no liquid electrolyte, no polymer, nor any other organic material present in thin film, solid state batteries; therefore, side reactions between the electrode and electrolyte are minimized. These characteristics make thin film batteries an ideal power source for devices operating in harsh environments including environments in which high gravity forces are encountered.

Although LiPON-based, thin film, solid state batteries have excellent specific-energy and energy-density characteristics when only the active materials are considered; these active layers have to be deposited on a thick substrate (i.e., normally thicker than 25 μm). In addition, these batteries are normally sealed with a thick polymer/metal laminate. Therefore, their practical energy densities are significantly reduced. Whereas, the energy density of the active material (cathode/LiPON/anode) can be as high as $\sim 1500 \text{ Wh/L}$, the addition of the substrate and the packaging to the battery reduces the energy density of the battery significantly. After adding a 50 μm thick metal substrate and a 0.1 mm thick packaging material, the energy densities of the batteries decreased to $\sim 229 \text{ Wh/L}$ and $\sim 71 \text{ Wh/L}$. Several methods can be used to increase the energy density of the battery, such as reducing the metal substrate thickness, replacing the metal substrate with a polymer or other low-density materials, increasing cathode thickness, reducing package thickness, etc.

One of the main differences among the manufacturers of thin films batteries is their selections of the substrates. In fact, the substrate properties selected by different companies often determine the technical route and market orientation of these thin film manufacturers. Front Edge Technology, Inc. (FET)^c uses mica as the substrate. The advantages of mica as a substrate are its light weight (its density is 1.4 g/cm^3) and its mechanical strength. Infinite Power Solutions, Inc. (IPS)^d used an ultra-thin (i.e., 10 μm to 20 μm thick) Cu-Ni alloy as a substrate. Another substrate has been used to vacuum seal the top of the device with a very thin polymer/epoxy. The total thickness of the device is $\sim 0.17 \text{ mm}$. Cymbet Corporation has developed a product named EnerChip^e, which is an on-chip, thin film battery with a configuration of $\text{Sn}_3\text{N}_4/\text{LiPON/LiCoO}_2$. The main advantages of the EnerChip battery are that it can be deposited directly on a silicon wafer and, along with other components in the electronic circuit, can go through the solder-reflow process. However, the energy density of this battery is about half that of Li batteries (i.e., Li/LiPON/LiCoO₂). Excellatron Solid State^f, LLC has developed a low-temperature deposition process for thin film batteries that enables the use of a flexible polymer (25 μm to 75 μm Kapton) or metal foil (25 μm to 50 μm Ni foil) as the substrate, thereby largely reducing the production cost of the thin film batteries. Planar Energy Devices (PED)^g used a “buried anode” of “reversed-structured,” thin film technology developed at the National Renewable Energy Laboratory. Recently, PED also reported the development of new-generation inorganic solid state electrolyte and electrode materials combined with a proprietary manufacturing process that overcomes the production and cost barriers to manufacturing low-cost, solid state, large format batteries. Several foreign companies and universities have been engaged in the development of thin film, solid state batteries. Most of these organizations are still using a LiPON-based, solid state electrolyte. GS NanoTech Co., Ltd.^h has developed a thin film electrolyte composed of $\text{Li}_{3.09}\text{BO}_{2.53}\text{N}_{0.52}$, which exhibits an ionic

^c www.frontedgetechnology.com/gen.htm

^d www.infinitepowersolutions.com/product

^e www.cymbet.com/content/products.asp

^f www.excellatron.com/advantage.htm

^g www.planarenergy.com/Technology.html

^h www.gsnanotech.co.kr/?z=contents.e_sub02_01_01

conductivity similar to the best values obtained for the LiPON-based electrolyte. These batteries have been used as the in-flight power source for electrical fuses in small ammunition[7].

Recently, the Massachusetts Institute of Technology (MIT)/ORNL/Draper Laboratory team[8] developed a high energy density miniature battery based on a thick cathode (LiCoO_2). These batteries use a thick, porous, high-density oxide cathode produced by sintering. Typical cathodes based on LiCoO_2 are 0.26 mm to 0.80 mm thick and have a sintered density of 74 to 87 percent. The electrode was prepared by repeated co-extrusion of a mixture of LiCoO_2 powder and aligned carbon fibers. When sintered at high temperature (i.e., $>700^\circ\text{C}$), the carbon fibers burn out, and aligned channels through which liquid electrolyte can be transported are formed. This structure exhibits a high electronic conductivity even in the absence of conductive additives. Adequate electrolyte-phase conductivity results from the low tortuosity of the available porosity when compared to conventional calendared electrodes. An energy density of more than 675 Wh/L at a C/3 discharge rate has been reported in a micro-battery with dimensions of 3 mm x 3 mm x 0.7 mm. This is the best energy-density value reported for miniature batteries. Furthermore, several research groups are in the early stages of demonstrating that the design and fabrication of three-dimensional, multifunctional architectures from appropriate nanoscale building blocks afford opportunities to improve energy storage for power/size scales that range from the nanowatt to microwatt power levels. For example, Hammond et al.[9] from MIT developed a way to create micro-batteries using a virus-based assembly method. The MIT team altered the virus' genes so it makes protein coats that collect molecules of cobalt oxide to form ultrathin wires. The resulting electrode arrays exhibit full electrochemical functionality. Further development of this technology will result in micro or nano power sources that can provide local power to activate molecules or power a single cell.

Another approach that is promising for long-term, low-power operations are betavoltaic batteries.

These batteries can generate electrical current using energy from a radioactive source that emits beta particles (i.e., electrons). The common radioactive sources used in betavoltaic batteries include ^{85}Kr , ^{147}Pm , and ^3H (tritium), which can convert beta particles to e-h pairs. Unlike most nuclear power sources that use nuclear radiation to generate heat, which then is used to generate electricity via thermoelectric and thermionic processes, betavoltaic batteries use a nonthermal conversion process. Betavoltaic cells have the potential to extend micro-scale battery life from hours to years, thus making possible a range of new applications such as wireless sensor arrays. City Labs, Inc.¹ has developed betavoltaic batteries based on tritium. The battery mainly consists of a beta source and a p-n junction that is used to absorb the electron and convert the radiation into useable electrical energy. The dimensions of the battery range from a few millimeters to a few centimeters. The battery is rated at an initial power level of ~ 75 nW (Model P100 battery) and it can last for 12 years, which is the half life of tritium. Widetronix, Inc.² improved the efficiency of the betavoltaic battery by incorporating silicon carbide, which exhibits low leakage current and less backscattering. These batteries are robust over a wide temperature range and can be powered down instantly. The life time of the battery varies from 2.6 years (^{147}Pm) to ~ 100 years (^{63}Ni), depending on the half life of the isotope selected. One device produced by Widetronix Inc. has a 6 mm x 6 mm footprint, a current density of 11 nA, and a power density of $2 \mu\text{W}/\text{cm}^2$.

Miniature power sources other than batteries also have been investigated with limited success. For example, Chmiola et al.[10] developed micro-supercapacitors by etching super-capacitor electrodes into conductive titanium carbide substrates. The volumetric capacities of these supercapacitors exceed, by a factor of two, those of micro- and macro-scale supercapacitors reported thus far. Significant progress also has been made recently on the development of micro-scale solid oxide fuel cells. Baertsch et al. at MIT reported solid-oxide fuel cells

¹ www.citylabs.net/index.php?option=com_content&view=article&id=10&Itemid=20

² www.widetronix.com/technology.php

with side dimensions of $\sim 525 \mu\text{m}$. Although these thin film fuel cells are operational, no energy density data were reported. Moghaddam et al.[11] created the world's smallest working hydrogen fuel cell, with dimensions of $3 \text{ mm} \times 3 \text{ mm} \times 1 \text{ mm}$. Pichonat et al.[12] developed low-cost miniature fuel cells based on proton-conducting porous silicon membranes; a power density of 17 mW/cm^2 was obtained for these devices. Most of the work on these miniature fuel cells is at the proof-of-concept stage. The efficiencies and energy densities of these devices still are not adequate for practical applications.

State of the art in miniature energy harvesting systems

Although a high-energy-density battery can extend operation time, electronic devices operated for long periods of time (i.e., >1 year) need to have rechargeable energy storage systems so the net volume of the energy storage devices can be minimized. Ideally, energy harvested or scavenged from the surrounding environment would be used to charge miniature energy storage devices. Possible energy sources to be harvested include solar, heat, vibration, radio-frequency (RF) radiation, and humidity.

Perhaps the best approach for energy harvesting is to use solar cells to convert light to electricity. The best multi-junction solar cells have an efficiency of ~ 42 percent, and the highest reported efficiency for thin film solar cells is 19.9 percent. This thin film solar cell, which was developed by the National Renewable Energy Laboratory^k, is based on copper-indium-gallium-selenide thin films, also known as CIGS. However, the efficiency of a solar cell strongly depends on the intensity of the light that falls on the cell. For example, the energy efficiency of a solar cell will decrease from $\sim 100 \text{ mW/cm}^2$ in direct sunlight to $100 \mu\text{W/cm}^2$ in an illuminated indoor environment. Other natural sources of harvested energy have an energy efficiency of less than 1 mW/cm^2 . Energy conversion from human behavior, such as a push button, a hand generator, heel strikes, etc., may generate a large amount of energy, but

they all require the movement of a human body and are much less suited for use as continuous energy sources.

Vibration energy can be harvested in several ways, including piezoelectric, electrostatic, and electromagnetic approaches. The energy efficiency of piezoelectric conversion devices ranges from $\sim 35 \text{ mJ/cm}^3$ to 350 mJ/cm^3 , while electrostatic devices have an efficiency of $\sim 4\text{--}40 \text{ mJ/cm}^3$ and electromagnetic devices have an efficiency of ~ 25 to 400 mJ/cm^3 . Seeman et al.[13] developed a miniature piezoelectric device that can provide 10 mV to 1 V to a MEMS device. The device response can be tuned to anticipate source frequency to achieve maximum power output. Fang et al.[14] developed a piezoelectric device consisting of a composite cantilever with a nickel metal mass. The device, which has a volume of 1.2 mm^3 , is expected to resonantly operate in a low-frequency environmental vibration by tailoring the structure dimension. The smallest piezoelectric device (i.e., 0.2 mm^3 volume and $1 \mu\text{W}$ power) was reported by Jeon et al.[15]. The smallest inductive energy-conversion device was prepared by Williams et al.[16] with a SmCo magnet, a polyimide membrane, and a planar gold-coil, GaAs substrate. The device has a volume of 4 mm^3 and a power of $0.3 \mu\text{W}$. The smallest capacitive-energy conversion device, reported by Ma et al.[17], has a volume of 13 mm^3 and a power of $0.065 \mu\text{W}$. Comparison of the state of the art in these three devices shows that the piezoelectric micro-device has the largest power density (i.e., $1.8 \mu\text{W/mm}^3$).

Thermoelectric generators are all solid state devices that convert heat into electricity[18]. One advantage of thermoelectric energy-harvesting devices is that they contain no moving parts so they are completely silent. Such generators have been used reliably for over 30 years of maintenance-free operation in deep space probes such as the National Aeronautics and Space Administration's Voyager missions. Several companies have developed miniature thermoelectric energy harvesting devices. The miniature thermoelectric device developed by

^k www.nrel.gov/solar/

^l www.poweredbythermolife.com/thermolife.htm

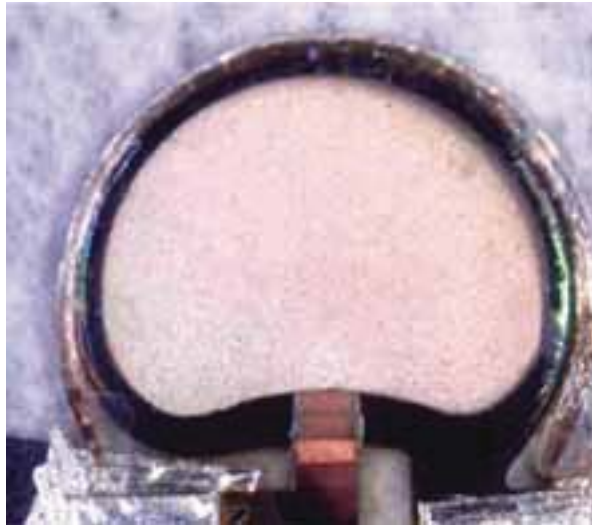


Figure 1: A thin film battery prepared by Front Edge Technologies. The sample used a 10-micron-thick mica substrate protected by a multilayer barrier deposited at PNNL. The total thickness of the sample is ~ 0.03 mm and the diameter of the device is < 10 mm.

Thermo Life Energy Corporation¹ has a diameter of ~ 10 mm. It can generate ~ 5 V at a temperature difference of 5 K or 11 V at a temperature difference of 10°C . In a typical miniature thermoelectric device, thousands of miniature thermoelectric couples are assembled, serially connected on a long strip, and then wound like tape with thermal contact on the tops and bottoms of the devices. The typical device made by Thermo Life Energy Corporation is smaller than a US penny.

Unlike the light, vibration, and thermal-gradient energy sources, RF power from radio stations, cell phone towers, etc., is an ideal candidate as a ubiquitous power source that is available most of time. The main disadvantage of RF power sources is the “fast fade” that occurs as the distance between an emitter station and a receiver increases. In the far field, the power density of an electromagnetic wave is proportional to the inverse of the square of the distance from the source. For example, the signal power strength of cell phones (RF/GSM 900 MHz) is $\sim 0.1 \mu\text{W}/\text{cm}^2$. The signal strength from a high-frequency source is even less (WiFi 2.4 GHz, $0.01 \mu\text{W}/\text{cm}^2$). As a comparison, the thermal electrical power generated from temperature gradients in a human body can be $25 \mu\text{W}/\text{cm}^2$.

Trends in thin film/miniature power sources

Power systems represent a significant challenge to the widespread deployment of wireless sensor networks and other microelectronic devices[13]. For example, aging aircraft need many sensors to monitor their operational health. Some aircraft sensors might need to be located in places that are inaccessible, making retrofitting with electrical power wires

difficult. In addition, electromagnetic signals from an aircraft need to be minimized to lower the possibility of detection and intercept. Ideally, these sensors would be operated by separate, self-sustainable power sources. However, even with the aggressive power consumption target of $100 \mu\text{W}$ per sensor, current miniature battery technology cannot even provide a year of autonomous operation. Wireless systems have traditionally been designed to use batteries as their power sources. If these ubiquitous wireless sensor networks are to become a reality, alternative power sources must be employed. On the other hand, although it is possible to power electronic devices by using scavenged power directly, an energy storage system will be required to provide a stable supply of energy when harvested or scavenged energy is not available. In other words, neither energy harvesting/scavenging devices nor thin film/miniature energy storage devices can provide the stable energy source needed to operate microelectronic devices for long time periods. Alternative approaches are needed to satisfy the continuously increasing energy and power requirements of these devices.

One of the most significant trends in thin film/miniature power sources in recent years is the combination of energy harvesting/scavenging devices and thin film/miniature power sources. When thin film batteries were first developed, in the mid-1990s, they were proposed as the solution for the future of the power sources for many high-power applications. However, in most cases, the energy required by microelectronic devices is much larger than can be provided by the capacity of thin film/miniature batteries. Many applications, such as sensor networks, need more power to operate independently in isolated environments. Without a continuous power supply, the lifetime of these devices will be very limited; some potential applications are not practical now and would be prohibitively expensive to be implemented. Considering this market reality, most companies working on rechargeable thin film/miniature batteries have turned to the “uninterruptible power supply” by combining their thin film/miniature batteries with energy harvesting devices. Among these products, energy harvested by photovoltaic cells has been the most widely promoted approach. Several companies, including FET, IPS, Cymbet, and

Excellatron, have proposed the use of a combination of photovoltaic cells and thin film batteries. In most cases, off-the-shelf photovoltaic cells were used, and the area and weight of photovoltaic cells were much larger than those of thin film batteries. Therefore, these combined devices were mainly used for demonstrations or to prove the concept.

Recently, a team led by FET (including Sandia National Laboratory, Pacific Northwest National Laboratory [PNNL], and the University of California at Los Angeles) have developed the smallest energy-source device that consists of a photovoltaic cell/thin film battery assembly. This technology won a 2010 R&D 100 Award. Several procedures have been used by FET and their partners to increase the practical energy density of thin film batteries. Figure 1 shows the thin film battery prepared by FET. The sample used a 10-micron-thick mica substrate and was protected by a multilayer barrier (3 μm to 5 μm thick) deposited by PNNL. PNNL's multilayer moisture barrier is composed of alternating layers of organic and inorganic materials deposited under vacuum. These multilayer barriers exhibit a very low water vapor transmission rate of $<10^{-6}$ g/m²/day. Ultra-thin batteries such as shown in Figure 1 were combined with an ultra-thin photovoltaic device (~ 10 μm thick and containing 10 single-junction, Si-based cells) developed by Sandia National Laboratory. Multiple thin film batteries and photovoltaic devices have been combined, and the entire system is environmentally sealed using a specially developed polymer coating. The volume of the system is approximately 1 mm³. It has an energy density greater than 300 Wh/L, which is the best energy density that has been achieved for a self-sustainable, miniature power source.

In addition to thin film batteries, improvements in the conductivity of solid state electrolytes will make high-capacity, high-energy-density, solid state batteries a reality within the next decade. One of the most promising solid state electrolytes is a sulfur-based material such as the $x\text{Li}_2\text{S} \cdot (100-x)\text{P}_2\text{S}_5$ system produced by melting and quenching methods[19]. This electrolyte exhibits a conductivity of 3×10^{-3} S/cm; however, its one disadvantage is its moisture sensitivity. In contrast, the air-stable, solid state electrolyte developed by Fu et al.[20] is promising. The glass produced by Fu et al. is composed of

$\text{LiTi}_2(\text{PO}_4)_3\text{-AlPO}_4$ (lithium titanium phosphate or LTP) and has a conductivity of $\sim 2 \times 10^{-4}$ S/cm. Its feasibility as a solid state electrolyte has been verified when used in Li-air batteries, Li-ion batteries, or other ionic devices. Because its ionic conductivity is relatively lower when compared with liquid or polymer electrolytes, the sheet has to be very thin to achieve reasonable current densities. At its typical thickness, which is about 0.15 mm, the LTP sheet is very fragile, and consequently, it is difficult to handle. Significant work is underway to produce a more flexible sheet with a similar conductivity and impermeability to water. For example, Polyplus Battery Company and Corning Incorporated were funded recently by the Defense Advanced Research Project Agency to develop a flexible LTP sheet for Li-air battery applications. We expect that a glass electrolyte that (1) is stable in air and water and (2) has the desired flexibility can be developed during the next three to five years. The conductivity of these glasses can be increased further to about 5×10^{-3} S/cm within the next 10 years.

Expected trends in the development of thin film/minature power sources within the next 15 years are summarized in Figure 2. The estimated performance is based on a device dimension (which includes all of the components, including package) of less than 1 mm or a thickness of less

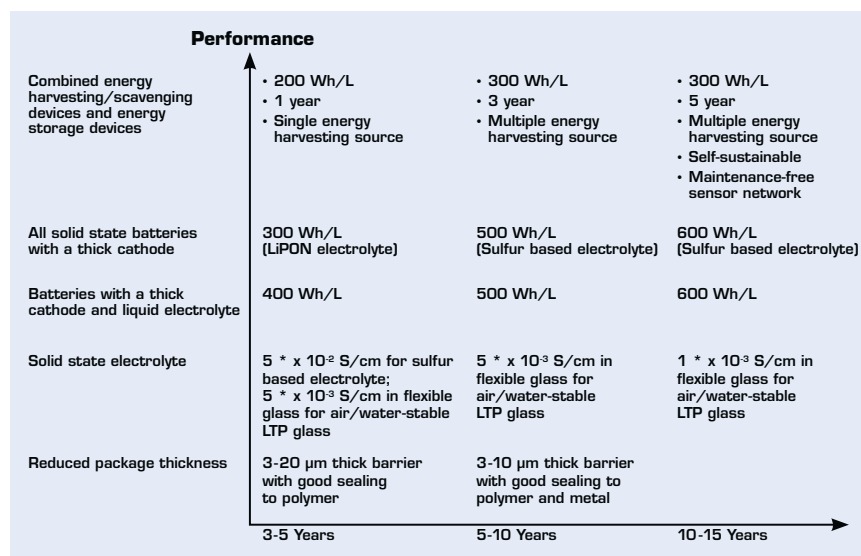



Figure 2: Summary of the technical forecast for thin film/minature power sources within the next 15 years. The performance estimate is based on a device dimension of <1 mm³ or a thickness of <0.3 mm.

than 0.3 mm. In the first three to five years, the main improvements will be in the individual components, including reduced package thickness, improved energy density of the electrode materials and cathode design, improved ionic conductivities, etc. Further developments within the next five to 10 years will lead to more stable power sources that can operate in more complicated environments by converting energy from different sources in the surrounding environment. Progress in the next 10 years will establish a solid foundation for future developments. We expect that a reliable ubiquitous power source with multiple energy harvesting capabilities and a total energy density of 300 Wh/L (with a dimension of less than 1 mm³ or a thickness of less than 0.3 mm) can be developed within the next 15 years. These devices will be the ideal power source for a self-sustainable, maintenance-free sensor networks and other microelectronic or MEMS systems.

About Pacific Northwest National Laboratory

Pacific Northwest National Laboratory is a US Department of Energy national laboratory where interdisciplinary teams advance science and technology and deliver solutions to America's most intractable problems in energy, the environment, and national security. PNNL provides world-renowned scientists and engineers, facilities, and unique scientific equipment to strengthen US scientific foundations and advance scientific discovery through innovation. 

References

- [1] Amatucci GG, Takeuchi ES, Kamat P, Barnett A, Snyder J, O'Handley RC, Abraham KM, Kohl P, Rolison DR, Blomgren G. Report of a National Science Foundation-Intelligence Community Workshop [Internet]. NSF-IC Power Sources Workshop; 24-25 April 2007; Chantilly (VA). National Science Foundation and the United States Intelligence Community; [undated webpage; accessed June 2010]. Available from: http://coewww.rutgers.edu/www5/nsfic/NSF_IC_MicroPower_Final.pdf
- [2] Excerpts from a conversation with Gordon Moore [video transcript, Internet]. Intel Corporation; 2005 [undated webpage; accessed June 2010]. Available from: ftp://download.intel.com/museum/Moores_Law/Video-Transcripts/Excepts_A_Conversation_with_Gordon_Moore.pdf
- [3] Bates JB, Dudney NJ, Gruzalski GR, Zuhr RA, Choudhury A, Luck CF, Robertson JD. Fabrication and characterization of amorphous lithium electrolyte thin and rechargeable thin film batteries. *Journal of Power Sources*. 1993;43(1-3):103-110.
- [4] Bates JB, Dudney NJ, Lubben DC, Gruzalski GR, Kwak BS, Yu X, Zuhr RA. Thin-film rechargeable lithium batteries. *Journal of Power Sources*. 1994;54(1):58-62.
- [5] Dudney NJ, Bates JB, Lubben D. Thin-film rechargeable lithium batteries. In: Kumta N, Rohrer GS, Balachandran U, editors. *Role of Ceramics in Advanced Electrochemical Systems*. Westerville (OH): American Ceramic Society; 1996. p. 113. (Ceramic Transactions; vol 65).
- [6] Dudney NJ. Solid-state thin-film rechargeable batteries. *Materials Science and Engineering B*. 2005;116(3):245-249.
- [7] Ji H, Kang SH, Cho SB. Development of miniaturized solid-state thin-film battery applied to electrical fuses at low temperature. In: *Proceedings of the 44th Power Sources Conference*; 14-17 June 2010; Las Vegas (NV). p. 281.
- [8] Lai W, Erdonmez CK, Marinis TF, Bjune CK, Dudney NJ, Xu F, Wartena R, Chiang Y-M. 2010. Ultrahigh-energy-density microbatteries enabled by new electrode architecture and micropackaging design. *Adv. Mater.* 2010;22(20):E139-E144.
- [9] MIT engineers work toward cell-sized batteries. MIT News [Internet]. 20 Aug 2008 [accessed June 2010]. Available from: <http://web.mit.edu/newsoffice/2008/virus-battery-0820.html>
- [10] Chmiola J, Largeot C, Taberna P-L, Simon P, Gogotsi Y. Monolithic carbide-derived carbon films for micro-supercapacitors. *Science*. 2010;328(5977):480-483.
- [11] Barras C. World's smallest fuel cell promises greener gadgets. *New Scientist* [Internet]. 07 Jan 2009 [accessed June 2010] Available from: <http://www.newscientist.com/article/dn16370-worlds-smallest-fuel-cell-promises-greener-gadgets.html>

[12] Pichonat T, Gauthier-Manuel B. Real-ization of porous silicon based miniature fuel cells. Journal of Power Sources. 2006;154(1):198-201.

[13] Seeman M, Sanders S. Harvesting micro-energy [slides, Internet]. CMOS Emerging Technologies Workshop; 18-20 Feb 2009; British Columbia, Canada. [last updated 20 Feb 2009; accessed June 2010]. Available from: <http://www.cmoset.com/uploads/11.5.B.pdf>

[14] Fang H-B, Liu J-Q, Xu Z-Y, Dong L, Wang L, Chen D, Cai B-C, Liu Y. Fabrication and performance of MEMS-based piezoelectric power generator for vibration energy harvesting. Microelectronics Journal. 2006;37(11):1280-1284.

[15] Jeon YB, Sood R, Jeong J-H, Kim S-G. MEMS power generator with transverse mode thin film PZT. Sensors and Actuators A: Physical. 2005;122(1):16-22.

[16] Williams CB, Shearwood C, Harradine MA, Mellor PH, Birch TS, Yates RB. Development of an electromagnetic micro-generator. IEE Proceedings - Circuits, Devices, and Systems. 2001;148(6):337-342.

[17] Ma W, Zhu R, Rufer L, Zohar Y, Wong M. An integrated floating-electrode electric microgenerator. Journal of Microelectromechanical Systems. 2007;16(1):29-37.

[18] Snyder GJ. Small thermoelectric generators. The Electrochemical Society's Interface. 2008;17(3):54-56.

[19] Hayashi A, Hama S, Morimoto H, Tatsumisago M, Minami T. Preparation of Li₂S-P₂S₅ amorphous solid electrolytes by mechanical milling. Journal of the American Ceramic Society. 2001;84(2):477-479.

[20] Fu J. Fast Li⁺ Ion conduction in Li₂O-Al₂O₃-TiO₂-SiO₂-P₂O₅ glass-ceramics. Journal of the American Ceramic Society. 1997;80(7):1901-1903.

Referenced Elements		Referenced Compounds	
Element Abbreviation	Element Name	Compound Formula	Compound Name
Al	aluminum	AlPO ₄	Aluminum Phosphate
As	arsenic		
B	boron	GaAs	Gallium Arsenide
Co	cobalt	LiCoO ₂	Lithium Cobalt Dioxide
Cu	copper		
Ga	gallium	Li-MnO ₂	Lithium Manganese Dioxide
H	hydrogen		
In	indium	LiPON	Lithium Phosphorus Oxynitride
Kr	krypton		
Li	lithium	LiTi ₂ (PO ₄) ₃	Lithium Titanium Phosphate [LTP]
Mn	manganese	SmCo	Samarium Cobalt
N	nitrogen		
Ni	nickel	Sn ₃ N ₄	Tin Nitride
O	oxygen		
P	phosphorus	Zn-MnO ₂	Zinc Manganese Dioxide
Pm	promethium		
S	sulfur		
Si	silicon		
Sm	samarium		
Sn	tin		
Ti	titanium		
Zn	zinc		

